No. 1 (New Series No. 197) January 198

12

# ANALYSIS

Edited by
CHRISTOPHER KIRWAN

ुस्तकालय

CONTENTS

um
xample of a language with classical logic for which bivalence cannot sumed — ALLEN HAZEN
nce meaning irreducible to use even within a use-theoretical ework — EIKE V. SAVIGNY
Fregean argument for the distinctness of sense and reference — R. M. SAINSBURY
ctive probability and indifference — ROY A. SORENSEN
h's defence of Lewy — A. J. DALE
bridge properties and defunct individuals: a reply to C. J. F. Williams — BARRY MILLER
nce — WILLIAM CHARLTON
lation and belief again — ARNOLD CUSMARIU
ske on knowledge and information — J. CHRISTOPHER MALONEY
king Nozick's sceptic: a better method — RAYMOND MARTIN
toxin puzzle — GREGORY S. KAVKA
and the Indians — MARTIN HOLLIS
e on ancient attitudes towards slavery — C. C. W. TAYLOR
g commits one to ethical beliefs — HARRY J. GENSLER
on universal prescriptivism and utilitarianism — INGMAR PERSSON
on conflicting moral obligations — ROBERT W. HOAG
an reproduction: a self-defeating strategy — LYNNE M. BROUGHTON
tors and democracies — LESLIE GREEN
e's scepticism concerning reports of miracles — ROY A. SORENSEN
existence and the cosmological argument — WILLIAM L. ROWE
en on the free will defence — DAVID GORDON

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003-2638

# *The Philosophical Quarterly*

NEW FROM BASIL BLACKWELL

Basil Blackwell Publisher is proud to announce that from January 1983 we will be publishing *The Philosophical Quarterly*, a welcome addition to an already established list of philosophy journals.

## Volume 33 No. 130 January 1983

| | |
|---|---|
| Rawls on Natural Inequality | MICHAEL GORR |
| Understanding Names | PETER CARRUTHERS |
| Two Worries about Mixing One's Labour | JEREMY WALDRON |
| Hegel on Property and Personality | DUDLEY KNOWLES |

## DISCUSSIONS

| | |
|---|---|
| Defining the Least Advantaged | ROY C. WEATHERFORD |
| The Intelligibility of Massive Error | BRUCE VERMAZEN |

## CRITICAL STUDIES

| | |
|---|---|
| A Tribute to Strawson (Zak van Straaten (ed) *(Philosophical Subjects: Essays presented to P. F. Strawson)* | JANE HEAL |
| Legitimacy and the Gadfly Challenge (Bruce A. Ackerman *Social Justice in the Liberal State)* | ANTHONY FLEW |
| BOOK REVIEWS | BOOKS RECEIVED |

**Medical Ethics Special Issue 1983**
This issue will be devoted to philosophical discussion of first-order moral problems such as abortion, euthanasia, the treatment of mental illness and the provision of medical care.

**Frege Special Issue 1984**
Contributions should be submitted by 31 December 1983 to: The Executive Editor, The Philosophical Quarterly, The University of St Andrews, Scotland KY16 9AL.

**Subscription Rates 1983**
Individuals: £9.00 (UK) £9.00 (overseas) $22.00 (USA)
$27.00 (Canada)
Institutions: £20.00 (UK) £20.00 (overseas) $45.00 (USA)
$55.00 (Canada

Orders, with remittance, or enquiries to: Sue Dommett, Journals Department, Basil Blackwell Publisher, 108 Cowley Road, Oxford OX4 1JF, England.

**Basil Blackwell Publisher · Oxford**

3589

ANALYSIS 43.1 JANUARY 1983

## ERRATUM

It has been pointed out that the 'New Series' numbering of recent issues of ANALYSIS has gone astray. The error first occurred in the October 1981 issue (Volume 41 No. 4) which should have been numbered 192. All four issues of Volume 42 have been numbered one lower than they should have been. The error has been corrected with the current issue which is numbered 197 instead of 196.

---

# AN EXAMPLE OF A LANGUAGE WITH CLASSICAL LOGIC FOR WHICH BIVALENCE CANNOT BE ASSUMED

*By* ALLEN HAZEN

GÖDEL long ago showed that classical first-order logic and classical number theory could be interpreted in intuitionistic logic and number theory. With a bit of reinterpretation, therefore – reading existential quantifiers as negations of universal quantifiers, etc. – an intuitionist can accept any theorem proved in classical number theory. Thus for an intuitionist, the language of classical first-order arithmetic is an example of a language in which the formal rules of classical logic are used, but for which bivalence cannot be assumed. For someone who has not been convinced by intuitionistic or other anti-realist arguments that bivalence cannot be assumed for languages, like that of arithmetic, in which an infinite domain is quantified over, the example will not be convincing. Thus I think there is some interest in the following example, which depends on a somewhat speculative epistemological hypothesis (about the concepts that can possibly be grasped by beings like humans), but which I find plausible, and which is entirely independent of the intuitionist critique of classical mathematics.

Russell believed that every possible concept was located somewhere in the ramified type hierarchy. (For convenience, let us assume with Russell that only concepts true or false of every entity in some well-defined domain – concepts that could be expressed by a predicate in some *bivalent* language – merit the name, and use some other term for whatever is expressed by the predicates of non-bivalent languages.) Perhaps no one now believes this in quite the form Russell did, but something like it is not altogether implausible. Let us assume at the base of our ramified hierarchy quantification over whatever non-abstract entities there may be – physical objects, certainly, and also minds or sense data if there are any such things – and predicates expressing whatever empirical properties and relations of these entities we are able to grasp. (This will include both those 'observational' properties whose presence in an object we are able to ascertain perceptually and also those 'theoretical' properties various theories tell us objects can have.) Since Russell's logicistic confidence that all mathematical concepts were ultimately logical seems to have been misplaced, let us also include in the base quantification over certain abstract entities and predicates expressing their basic mathematical properties – quantified variables ranging over the natural numbers and predicates for successor, sum, and product, certainly; more contentiously, perhaps sets and membership too. Thus our zero-order variables will range over concrete objects and certain simple mathematical ones. Let us build a ramified hierarchy on this base – first-order variables will range over first-order-definable properties of the basic objects, second-order variables over properties definable by quantification over those first-order-definable properties, third-order variables over properties definable by quantification over *those* properties, and so on. Let us allow the ramified hierarchy to extend into the transfinite (Russell and Whitehead's argument for only recognizing finite orders is unconvincing). Then it is not implausible to assume that every humanly graspable concept – every concept any being like a human could possibly formulate – is located somewhere in this ramified hierarchy.

No single formal language can describe the whole of this hierarchy, but we may think of it as described by the union, as it were, of a hierarchy of formal languages having quantified variables of higher and higher transfinite orders. For any two such languages there will be a third, with quantified variables of at least as high an order as the highest-order variables in either of the two, into which both can be translated. (Indeed, if we are willing to rule out certain rather arbitrary kinds of restriction on the expressive power of individual languages, we may assume that they are linearly ordered: that of any two such languages, one will have higher-order variables, and the other will be translatable into it.) In each of these languages, full classical quantificational logic holds for each order of quantified variables, for these are all bivalent languages.

For convenience, let us assume that the languages have their quantificational machinery formulated so that (as in typical systems of natural deduction) there is a set of free variables or parameters or dummy constants of each order for which there are quantified variables, and so that, whenever an existentially quantified formula is given, an instance of it, using one of the free variables, etc., of appropriate order, may be inferred (at least in a subordinate proof, or with a flag, etc.).

Now, it is a general metamathematical fact about classical quantificational logic that, given any language with classical propositional logic and any class of expressions of that language, it is possible to introduce a new sort of bound variables having the expressions of that class as substituends, and to use full classical quantifier logic in connection with the new variables, without making anything in the original language provable that wasn't provable before. (The language with the new sort of quantified variables is a *conservative extension* of the old language; any theory formulated in the new language with none of the new bound variables occurring in its axioms will be a conservative extension of the theory in the original language having the same axioms.)

Given one of the languages in our hierarchy, take the free variables or parameters of every order as a substitution class, and introduce what we may call *general* variables having this substitution class. (We may intuitively think of the general variables as ranging over the union of the domains the various orders of variables in the original language range over.) We may, by the theorem cited, use classical quantifier logic with the general variables. The logical relations between the original language and the language with the general variables are easily formulated. Take any condition formulable in the language (i.e. any sentence with a blank that could grammatically be filled by a free variable or parameter). Call the sentence that results from filling the blank with a variable and binding that variable with an initial existential (universal) quantifier its existential (universal) completion with that variable. Then the existential completion of a condition with a general variable follows from its existential completion with an 'original' variable of any order; its universal completion with an original variable of any order follows from its universal completion with a general variable.

Define the *general theory induced* by a language in our hierarchy as the set of all sentences with quantified general variables implied by the set of all true sentences of the original language. Then we have the result that the general theory induced by one language is always consistent with the general theory induced by (and hence with all the truths of) any other language in the hierarchy. Proof: there is a third language in the hierarchy into which both original languages can be translated. The general theories induced by the two languages will be translatable into subsets of the general theory induced by the third language. Ignoring the possible need for translation that arises from the fact that the notation for, e.g., the orders of variables in one language may not be compatible with that in another, we may generalize and say that anything formulated with general variables implied by truths of one language will be implied by truths of any language in the hierarchy of sufficient expressive power. Thus we may think of a sentence with general variables we derive from assumed truths of some ramified language of our hierarchy, not as saying something somehow specific to that particular language, but as belonging to a single 'theory' induced by the whole hierarchy. Thus, when we use an extension of one of our languages with general variables, we may think of the general variables not as ranging simply over the union of the domains of the various orders of 'original' variables in the language, but as ranging over all the concepts defined at any order of the whole ramified hierarchy. Thus it is possible for us to have a language which uses classical logic throughout (both with regard to its original, ordered, bound variables and with respect to its general variables) and, on our modified Russellian assumption, has variables that range over absolutely all humanly definable concepts.

If the assumption is true, however, the language cannot be assumed to be bivalent, for if it were, the conditions formulable in it with the help of general variables would themselves express humanly definable concepts. In that case there would be no avoiding an unrestricted principle of comprehension for concepts: for every condition formulable in the language there would be a concept holding of precisely the entities satisfying the condition. Since on the intended interpretation the general variables range over all concepts, this principle is expressed in the usual way by a schema in the language with general variables, and all the paradoxes that usually follow from the unrestricted comprehension schema can be derived from it. Thus, for example, there is a formula in the language, with two free general variables, saying that concept X holds of object (i.e. concept or base-level object) Y. Identifying the variables and negating, we get the condition that concept X does not hold of itself. To conclude, by the comprehension principle, that there is a concept holding of just those concepts fulfilling this condition is to derive Russell's paradox.

There is nothing mysterious or meaningless about the language with general variables. All that is needed to establish its legitimacy, given our modified Russellian assumption, is a standard metamathematical result about classical predicate logic, and the principles relating quantifications of general variables to quantifications of typed variables give it a natural (though not, of course, a complete) proof procedure. Now, we find it natural to assume that languages we understand are bivalent (or could at the very least be made bivalent by tightening up the definitions of vague predicates). Intuitionism challenges this natural assumption for, e.g., the language of elementary number theory, but does so by adopting a constructive notion of truth. What I have tried to argue is that a consideration of Russell's explanation of the semantic paradoxes suggests an analogous challenge which cannot be avoided simply by denying, as a Realist, that truth and proof are as intimately related as the intuitionists hold. The assumption of bivalence for a language with variables ranging over all concepts leads to contradictions independently of any assumption that the truths of that language — or of the languages in the ramified hierarchy — are all provable.

The non-bivalent languages may, on both positions, be given a classical logic, but at the expense of obscuring distinctions. In the example of a language with general variables, the distinctions that could be made with a more sensitive group of logical constants might help in the formulation of an adequate comprehension principle for concepts.

*460 Riverside Drive*
*New York, NY 10027, U.S.A.*

© ALLEN HAZEN 1983

## REFERENCES

Gödel's result is in 'Zur intuitionistischen Arithmetik und Zahlentheorie', *Ergebn. math. Colloq.* 4 (1933), translated in *The Undecidable*, ed. Martin Davis. The locus classicus for Russell's views is 'Mathematical Logic as based on the Theory of Types', *Am. Jour. Math.* 30 (1908), reprinted in *Logic and Knowledge*, ed. Robert C. Marsh, and *From Frege to Gödel*, ed. Jean van Heijenoort. Transfinite ramified type theory is discussed in the last part of Hao Wang's *Survey of Mathematical Logic*, reprinted under the title *Logic, Computers, and Sets.* The result about adding classical quantifier logic with a new sort of variables to an arbitrary language with classical propositional logic can be extracted from section 32 of Church's *Introduction to Mathematical Logic*; it is an almost immediate consequence of Gentzen's Hauptsatz. Its semantic implications are interestingly discussed in Hacking, 'What is Logic?' *Journal of Philosophy* 76 (1979) and Göran Sundholm's comments on it, *Journal of Philosophy* 78 (1981). For reasons to suspect that intuitionistic logic is an appropriate 'more sensitive' logic for the language with general variables, see the present writer's 'On a possible misunderstanding of Kripke's Semantics for Intuitionistic Logic', *Analysis* 42.3, June 1982, pp. 128–133.

# SENTENCE MEANING IRREDUCIBLE TO USE EVEN WITHIN A USE-THEORETICAL FRAMEWORK

*By* EIKE V. SAVIGNY

## I

IMAGINE an ambitious linguist (or philosopher) coming up with a general semantics for languages in some canonical orthographic form:

> The meaning of a word is the number of its letters; the meaning of a sentence is the sum of the meanings of its component words.

Despite the obvious simplicity and elegance of his definition, the linguist fails to win our approval. However, if we tell him that we are not prepared to accept length for meaning we ought to give him reasons. On the other hand, if some linguist or philosopher offers the truth conditions of sentences as their meanings, he ought to give us reasons as well. What way is there to argue for classifying a given type of entities as sentence meanings?

The object of the present paper is to canvass a certain way of establishing that entities described as sentence meanings by a semantics really are sentence meanings. Such propaganda will amount to something like an explication of the concept of senten meaning in terms of something else, and usually reductive ente prises are motivated by a moderate degree of hostility towards the target. However, far from explaining sentence meaning away, I sha try to show why, even if we try to describe a language from a som what radical 'meaning is use' point of view, we have to accep sentence meanings as entities which are not reducible to the use the sentences they are meanings of.

The 'somewhat radical "meaning is use" point of view' holds tha a conventional language has been completely described as soon a we are given a set of rules which specifies all manners of producin each and every conventional utterance meaning. Therefore, th object of the description is to describe completely what speakers o the language in question can say, and how they can say it. Fo example, if by using a given sentence the speaker can, by virtue o the rules of his language, either warn the addressee not to approac the dog, or recommend him to buy the dog, or state his claim tha the dog be tied up, the description of his language has to specif under which circumstances he will say the first, the second, or th third. As is clear from the examples, I construe conventional utterance meaning within the framework of speech act theory. There i but one reason for doing so, *viz.* that speech act theory seems to b the only available theory which takes the idea seriously that there i utterance meaning which is governed by linguistic rules and which i not predictable from the uttered sentence alone (even if this sentence is unambiguous and free from indexical features). Any different concept of conventional utterance meaning satisfying a 'meaning is use' proponent would equally fit the bill; if no concept of conventional utterance meaning satisfies some 'meaning is use' opponent, let him be reminded that the point of the present enterprise is to account for the unavoidability of sentence meaning (which he hopefully accepts) even within a theory which he actively dislikes.

## II

Languages may in fact do without sentence meanings; there is, for instance, the extreme case of a language comprising but one sign, or several freely interchangeable signs. Rules which specify utterance meanings in such a language operate both on the fact that the sign (or one of them) is used and on further circumstances of the situation. Since differences between signs are irrelevant, the utterance meaning rules need not operate on features of signs; therefore, they need not operate on sign meanings in order to provide for every utterance meaning. Since nothing else is required, we need not posit sign meanings. To put it more harshly: we must not posit sign meanings in such a case because ascribing a meaning to a sign would be arbitrary – any other ascription would be just as justified.

The idea that 'meaning is use' has not, however, been invented for languages comprising only expressions which it is pointless to distinguish from each other. One might even argue that such languages are not real languages at all. Such arguments are likely to start from the premise that the serviceability of a language for saying indefinitely many things, and new things in particular, goes with its having a recursive syntax. But this is wrong for several reasons. First, the rules may specify new signs without a syntax, *viz.* if they specify general features that make patterns of behaviour iconic signs. Second, there may be new signs but no new things to say, as when arbitrarily many repetitions of one and the same element are all admitted as sentences of the language although it makes no difference for the utterance meaning which one the speaker uses (he may for instance decide on length according to expected difficulties in transmission). Third, there may be indefinitely many new things to say even where there is only one sign; for the utterance meaning rules can specify utterance meanings according to circumstances of the utterance situation in such a way that the variation of such circumstances regularly goes with variation of the utterance meaning.[1] However, we need not argue that no uninteresting cases exist in order simply to find them uninteresting in certain contexts. So let us admit that languages may do without sentence meanings and go on to inquire whether sentence meanings are avoidable where they *prima facie* play a crucial role. To have first considered more primitive languages will prove to be a good starting point for clarifying what this crucial role is.

## III

Let us consider a language (like ours) where it often matters which sign the speaker uses. That is to say, there are situations in which using sign $\sigma_1$ has an utterance meaning different from that of using sign $\sigma_2$ in the same situation. 'The same situation' has of course to be interpreted relative to the utterance meaning rules; so we had better say that there have to be utterance meaning rules which cannot give the required difference in utterance meaning while neglecting the fact that different signs are used. Since we require our description to account for all utterance meanings, its utterance meaning rules must, therefore, take into account which, or what kind of, sign is used.

'Which, or what kind of' makes a theoretical difference. For if the number of signs is finite, the utterance meaning rules may, in principle, be stated for each sign separately by pairing it off with the different kinds of circumstances in which it can be used to say different things. This is not possible for indefinitely many signs.

---

[1] See my 'Saying New Things in Finite Languages', *Philosophical Studies* 40, 1981, 259–267.

I think, however, that it can be shown that usual standards of intelligent theorizing may require us to have our utterance meaning rules operate on kinds of signs instead of enumerated lists even where there are only finitely many, or even few of them. To do this, I need to have recourse to an example which is taken from the language mentioned in footnote 1 but adapted here to everyday gesture language. 'x has a standard interest in y' is a clear-cut concept in the former, and the application to the latter may be taken to be concerned with cases where it is clear enough.

In considering situations where the direction gesture (moving one's arm in a certain direction) as well as the beckoning gesture can be used either to request the addressee to do something or to permit him to do something, we find that the following holds ($x$ uses the sign, $y$ is present):

| features of the situation<br>sign used | $x$ has a standard interest in $y$'s doing $V$ | $y$ has a standard interest in $x$ letting him do $V$ | further features of the situation |
|---|---|---|---|
| direction gesture | $x$ requests $y$ to do $V$ | $x$ permits $y$ to do $V$ | *that* $y$ *moves in the direction indicated* is suitable for $y$'s doing $V$ |
| beckoning gesture | | | *that* $y$ *approaches* $x$ is suitable for $y$'s doing $V$ |

This can encourage us to assign what is denoted by the italicized that-clauses to the signs (let us just call it 'the $m$' of the sign) and to formulate rules in the following manner:

(1) $x$ has a standard interest in $y$'s doing $V$ and uses a sign with an $m$ which is suitable for $y$'s doing $V$; he thereby requests $y$ to do $V$.

(2) $y$ has a standard interest in $x$'s letting him do $V$; $x$ uses a sign with an $m$ which is suitable for $y$'s doing $V$. He thereby permits $y$ to do $V$.

This move is far from necessary, of course; but it can be judged with respect to its theoretical fruitfulness according to whether it serves for further cases, i.e. with respect to further utterance meanings and further signs. If we find that we can state a rule

(3) $y$ has a standard interest in $x$'s letting him do $V$; $x$ uses a sign with an $m$ which is unsuitable for $y$'s doing $V$. He thereby insists upon $y$ omitting $V$,

and more and more similar rules; and if we find that these rules work equally well for utterance meanings produced by using the stopping gesture (palm towards the addressee) when we assign *that* *y does not approach x* to it, and for more and more such assignments to signs; then it would be *ad hoc* not to make use of such systematic regularities.

The $m$'s, of course, are good candidates for sentence meanings. Therefore, sentence meanings enter into our description of a language as soon as there are good theoretical reasons for having utterance meaning rules that operate on different kinds of signs. This condition is fulfilled if there are indefinitely many signs, but it may also be fulfilled for finitely many signs, and even if there are few of them.

## IV

In order to speak of *the* meaning of a given sentence, we have to require that there be a singular assignment; this requirement is included in the definition to be stated below by referring to assignments within the best description of the language in question. (If there is more than one best description, and they differ on sentence meaning assignments, then talk of 'the' meaning is illegitimate.) There may be different constraints on 'best' descriptions, corresponding to partially different concepts of sentence meaning. I will state only two which seem to be essential. First, the best theory has to be true in order to guarantee that new evidence can only change our views on meanings rather than the meanings. For imagine the best theory could be refuted by new evidence. Such a refutation might require us to regard some of the theoretical statements, 'the meaning of sentence $\sigma$ is $m$', as false. But $m$ is *by definition* the meaning of $\sigma$ if it is assigned to $\sigma$ by the best theory. Thus we must exclude the possibility that the best theory might be false. (This means, of course, that our views as to which theory is the best one and which sentence meanings are the real ones will forever remain provisional.) Second, the $m$ assigned to a sign must not be that sign's specification: for this would mean excluding synonymity of sentences by definition. The definition is to express the idea that sentence meanings are theoretical features that have to be attributed to the sentences in order to provide for the empirical correctness and completeness of a set of utterance meaning rules:

> Let $L$ be a language; let $S$ define, for every sentence $\sigma$ of $L$, $S(\sigma)$; let $P$ be a set of rules[2] $P_j$, each of which has the form: 'If any sentence $\sigma$ with $S(\sigma) = m$ is uttered in a situation $s$ such that $R_j(m, s)$, then uttering $\sigma$ in $s$ has the utterance meaning $(i_j(m,s), p_j(m,s))$.'[3] Then a sentence $\sigma$ of $L$ has sentence meaning $m$ in $L$ iff (1) $(P, S)$ is the best description of $L$, and (2) $S(\sigma) = m$.

---

[2] We may say instead that $P$ defines, for pairs of situations $s$ and sentence meanings $m$, the meanings of utterances in $s$ of sentences with $m$. I find the above formulation more informative because it stresses the point of the relation $R_j$ which is characteristic for each rule.

[3] Again, this terminology is speech-act theoretical, with '$i$' for the illocutionary force and '$p$' for the proposition (or rheme, with Austin). Both are functions of $m$ and $s$.

This definition of sentence meaning does not do the job of a semantics, of course; it states what a semantics $S$ must achieve in order to supply specifications of sentence meanings. I shall now put it to use in showing why the meaning of a sentence cannot be reduced to its use, although the concept of sentence meaning has been introduced in accordance with a 'meaning is use' point of view. This use I identify with illocutionary act potential for the simple reason that Alston's definition of sentence meaning in terms of illocutionary act potential[4] is the most straightforward meaning-is-use definition that I know of. Two prior points: I take the illocutionary act potential (iap) of a sentence to be its potential *vis à vis* whole utterance meanings rather than illocutionary forces (e.g. for requesting-that-you-come rather than for requesting); and I take 'potential' to be a function from occasions of use to utterance meanings rather than a class of utterance meanings. (I do not think there is another plausible interpretation.) The claim, then, is that the illocutionary act potential of a sentence so construed is not identical with its meaning as defined above.

## V

Let us consider a specific $\sigma_1$ such that $S(\sigma_1) = m$. For each pragmatic rule $P_j \in P$ we find that if $\sigma_1$ is uttered in some $s$, then either $R_j(m, s)$ holds and $(i_j(m, s), P_j(m, s))$ results as utterance meaning, or $R_j(m, s)$ does not hold, and no utterance meaning results from $P_j$. So for $\sigma_1$, $P$ pairs off situations of uttering $\sigma_1$ with utterance meanings. If $S(\sigma_2) = m$ too, then, if $\sigma_2$ is uttered in some $s$, $R_j(m, s)$ will hold precisely in the same cases where $\sigma_1$ is uttered; so occasions of uttering $\sigma_1$ and $\sigma_2$ are paired off with precisely the same utterance meanings. Therefore, if $S(\sigma_1) = S(\sigma_2)$, then iap $(\sigma_1)$ = iap $(\sigma_2)$.[5] (Note that $\sigma_1$ and $\sigma_2$ belong to one and the same language and that therefore the relevant $P$ is the same for both of them.)

If the meaning of a sentence is to be identified with its illocutionary act potential, then the converse must hold too. In line with the central idea of the above definition, its truth might thus be argued: 'If $\sigma_1$ and $\sigma_2$ have the same iap, then they need not be distinguished by a theory whose only aim is to generate utterance meanings correctly and completely. Of two such theories one that assigns different meanings to $\sigma_1$ and $\sigma_2$ is, *ceteris paribus*, less simple than one which makes them synonymous. The best theory, then, cannot assign different sentence meanings to sentences with the same iap.'

---

[4] W. P. Alston, *Philosophy of Language*, Englewood Cliffs, N.J., 1964, p. 36.

[5] If there are any two English sentences with, e.g., identical truth-conditions but different iap's according to the best $(P, S)$ for English, then their truth-conditions are not their sentence meanings in the sense of the above definition. This should be kept in mind because the definition will be presupposed in the following argument.

I think that this reasoning goes wrong because it neglects at least one reason for wanting 'sentence meaning' to be conceptually different from 'utterance meaning'. I shall comment on this substantial drawback, leaving aside questions as to how to compare gains in simplicity resulting from reduced numbers of meanings with possible losses resulting from introducing complicated meanings, or how to prove that for any theory assigning different meanings to sentences with the same iap, there really is a simpler theory assigning one and the same meaning to the sentences in question (a claim different from the claim that such a theory would be simpler).

My objection against the proposed biconditional is that it fails to account for the role which our concept of sentence meaning plays in everyday translation. (In fact I suspect translation to be the reason why we have the concept at all.) Take two languages $L_1$ and $L_2$ which have been described separately and in which signs $\sigma_{11}$, $\sigma_{12}$ and $\sigma_{21}$, $\sigma_{22}$ have meanings $m_{11}$, $m_{12}$ and $m_{21}$, $m_{22}$ respectively. Let $\sigma_{11}$ and $\sigma_{12}$ have the same $\text{iap}_1$ while $\sigma_{21}$ and $\sigma_{22}$ have $\text{iap}_{21} \neq \text{iap}_{22}$. Now it may well be the case that $m_{11}$ and $m_{21}$ happen to be identical, as are $m_{12}$ and $m_{22}$ (so that $\sigma_{11}$, $\sigma_{21}$ and $\sigma_{12}$, $\sigma_{22}$ are intertranslatable as a result of the semantics of $L_1$ and of $L_2$). This is possible if $P_{L_1} \neq P_{L_2}$. If $\text{iap}\,(\sigma_{11}) = \text{iap}\,(\sigma_{12})$ were sufficient for $m_{11} = m_{12}$, it would then follow that $m_{21} = m_{22}$, which contradicts $\text{iap}_{21} \neq \text{iap}_{22}$. So the fact that two sentences determining the same iap in one language may be translatable into two sentences determining different iap's in the target language is a sufficient reason for not identifying the first two sentences' meanings. (Even if the proposed biconditional held, translation arguments would show that $\text{iap}(\sigma_1) = \text{iap}(\sigma_2) \equiv S(\sigma_1) = S(\sigma_2)$ does not warrant Alston's $\text{iap}(\sigma) = S(\sigma)$. For it may be the case that $S_{L_1}(\sigma_1) = S_{L_2}(\sigma_2)$ while $\text{iap}_{L_1}(\sigma_1) \neq \text{iap}_{L_2}(\sigma_2)$, which is incompatible with $S_{L_1}(\sigma_1) = \text{iap}_{L_1}(\sigma_1)$ and $S_{L_2}(\sigma_2) = \text{iap}_{L_2}(\sigma_2)$. Compare 'He did not say it in so many words' and 'Er hat es nicht mit so vielen Worten gesagt' which, though possibly treated alike in semantic theories of English and German respectively, are used with different utterance meanings.)

It seems to emerge that reducing sentence meaning to illocutionary act potential would require us to describe all possible languages together as if they were one.[6]

*Fakultät für Geschichtswissenschaft*
*und Philosophie, Universität Bielefeld,*
*Postfach 8640, 4800 Bielefeld, W. Germany*

© EIKE V. SAVIGNY 1983

---

[6] I have to thank many friends and colleagues for critical comment on earlier versions of this paper. I am especially indebted to Jonathan Bennett whose encouraging comments helped me to substantially clarify my claims and correct my arguments; and to Andreas Kemmerling, Lorenz Krüger, and Bernhard Thöle from the University of Bielefeld.

# ON A FREGEAN ARGUMENT FOR THE DISTINCTNESS OF SENSE AND REFERENCE

*By* R. M. SAINSBURY

THE following argument could, without excessive violence to the text, be attributed to Frege (in the first two pages of 'On Sense and Reference'). For symbols, see note on page 14 below.

(1) $a = b$
(2) '$a = a$' is *a priori*
(3) '$a = b$' is not *a priori*
(4) if $s(e) = s(e')$, '$\ldots e \ldots$' is *a priori* iff '$\ldots e' \ldots$' is *a priori*. (In other words, if two expressions have the same sense, replacing one by the other will never affect *a priority*.)
(5) $\forall x \forall y$ (if $x^\frown$'='$^\frown y$ is true, then $r(x) = r(y)$). (In other words, the names in a true identity sentence have the same reference.)
(6) $\forall x(s(x) = r(x))$. (The hypothesis for refutation: there is no distinction between sense and reference for names.)
1, 5 (7) $r(\text{'a'}) = r(\text{'b'})$ from (1) and (5)
1, 5, 6 (8) $s(\text{'a'}) = s(\text{'b'})$ from (6) and (7)
1, 4, 5, 6 (9) '$a = a$' is *a priori* iff '$a = b$' is *a priori* from (4) and (8)
1, 2, 3, 4, 5, 6 (10) '$a = b$' is *a priori* and '$a = b$' is not *a priori* from (2), (3) and (9)
1, 2, 3, 4, 5 (11) not: $\forall x(s(x) = r(x))$.

Frege is generally taken to have desired to prove not merely (11) but rather the stronger

$$\forall x(s(x) \neq r(x)).$$

The first step to the required generalization is straightforward. One adds the premise that for any two co-referential names we have the analogue of (3). This gives the conclusion (upon the relevant premises) that any two co-referential names have senses distinct from their reference. The second, and more difficult step, to the generalization would use the idea that for any name one *could* introduce a co-referential one. Though the generalization presents some interesting issues, I shall not pursue it.

That (11) follows from (1), (2), (3), (4) and (5) is, I think, beyond question. But, instead of drawing this conclusion upon these premises, one might prefer to reject some assumption other than (6). I shall take it for granted that (1) and (5) hold. I shall also not challenge (2), though Frege himself should not have accepted it, since on his theory one can understand a sentence even if it contains an empty name (so one precondition for *a priority* is fulfilled);

but if it contains such a name the sentence will not be true, so there is no question of knowledge. My reason for not worrying about this is that we could relativize the predicate in (2) and (3), for example using 'is *a priori* relative to the knowledge that a and b exist', and then generalize (4) so that it holds for all relativizations. The upshot is that the only serious candidates for rejection, permitting the retention of (6), are (3) and (4).

Suppose we say:

> $\sigma$ is *a priori* iff anyone who understands $\sigma$ can, solely in virtue of this understanding, come to know the truth of the proposition which $\sigma$ expresses.

I now present, without endorsing, the case for one who would retain (6). '$a = a$' and '$a = b$' express the same proposition. (They say the same about the same.) Hence '$a = b$' is *a priori*, since one needs no more than an understanding of this sentence to be able to know the truth of the proposition it expresses, viz. that $a = a$. What gives rise to apparent absence of *a priority* is the possibility of not knowing, concerning a sentence one understands and a proposition which one knows, that the one expresses the other. The dislocation between *understanding* and *proposition* permits a gap between knowing a proposition which an understood sentence expresses and knowing that the sentence expresses this proposition. The consequence is that (3) is false.

To avoid this, one might prefer the following account of *a priority*:

> $\sigma$ is a *priori* iff anyone who understands $\sigma$ can, solely in virtue of this understanding, come to know that $\sigma$ expresses a true proposition.

This justifies (3): for even if one knows the truth (3) expresses, one may be in no position to know that it expresses this truth. But (4) is, on this account, false, as the very instance in question shows. Putting 'b' for one occurrence of 'a' in '$a = a$' transforms the sentence into one such that one cannot know, solely on the basis of understanding it, that it expresses a truth.

One obtains variants of the argument by replacing *a priority* by some other concept. For example, one might think that in place of (4) one should have

(4′) if $s(e) = s(e')$, '. . . e . . .' has the same content as '. . . e′ . . .';

and one might hold that this fixes what one means by 'content' and is therefore unrejectable. But then, in place of (2) and (3), one would need the highly controversial

(3′) '$a = a$' and '$a = b$' differ in content.

To show (3′) to be true one would require considerations already sufficient, independently of the present argument, to establish the distinction between sense and reference.

Or one might prefer, in place of (4),

(4″) if s(e) = s(e′) then 'A believes that . . . e . . .' is true iff 'A believes that . . . e′ . . .' is true.

This would return us to a situation parallel to the one produced by (4). The defender of (6) may accept that

(3″) 'A believes that a = a' is true iff 'A believes that a = b' is true

and use this to show the falsehood of (4″);[1] or he may accept (4″) and use this to deny (3″).

I do not assert (6) but claim only that it has not been persuasively argued against in the Fregean argument given. A persuasive argument along anything like these lines would have to dig deeper into such notions as *a priori*, *proposition* and *understanding*.

*Bedford College*
*(University of London),*
*Regent's Park, London*
*NW1 4NS*

© R. M. SAINSBURY 1983

## NOTE ON SYMBOLS

'a' abbreviates 'Hesperus', 'b' abbreviates 'Phosphorus'.

'e' is a variable ranging over expressions.

'x', 'y' are variables ranging over names.

'r' is a functor which takes a name of a name to form a name. Semantically, as (6) suggests, the name it forms names the reference, if anything, of what is named by what it operates on.

's' is a functor with the same syntax as 'r'. The name it forms names the sense, if any, of what is named by what it operates on.

'σ' is a schematic letter whose substitutes are sentences.

---

[1] David Wiggins denies something close to (4″) in his 'Frege's Problem of the Morning Star and the Evening Star', in Matthias Schirn (ed.), *Studies on Frege II: Logic and Philosophy of Language* (1976).

# SUBJECTIVE PROBABILITY AND INDIFFERENCE

*By* ROY A. SORENSEN

IN 'Is Epistemic Preferability Transitive?' (ANALYSIS 41.3, June 1981), I attack the following axiom appearing in Chisholm's and Kiem's 'A System of Epistemic Logic' (*Ratio*, December 1972):

$$\text{(A2)} \quad (\sim(pPq) \,\&\, \sim(qPr)) \supset \sim(pPr)$$

This axiom is read: If it is not the case that p is epistemically preferable to q, and it is not the case that q is epistemically preferable to r, then it is not the case that p is epistemically preferable to r. Peter Millican has pointed out that my argument against (A2) is unsound if epistemic preferability is given a probabilistic interpretation ('On the Transitivity of Epistemic Preferability', ANALYSIS 42.2, March 1982). Nicholas LaPara gives such an interpretation in 'Chisholm, Kiem, Preferability' (*Ratio*, June 1975). In the following two paragraphs I provide another counterexample to (A2) that also shows that epistemic preferability cannot be given a probabilistic interpretation.

Suppose there are 21 buckets of water $b_1, b_2, \ldots, b_{21}$ such that bucket $b_i$ contains water at i degrees centrigrade. Jim makes pairwise comparisons between $b_i$ and $b_{i+1}$ where $1 \leqslant i < 21$ by putting his hands in the appropriate buckets. Since Jim cannot discern a temperature difference of 1 degree by this method, he is epistemically indifferent between propositions $p_i$ and $p_{i+1}$ where '$p_i$' reads 'The bucket containing water of the highest temperature of all 21 buckets is bucket i'. So if the epistemic indifference relation is transitive, Jim should be indifferent between $p_1$ and $p_{21}$. However, Jim can discern a temperature difference of 20 degrees and so he is not indifferent between $p_1$ and $p_{21}$. Therefore, epistemic indifference is not transitive. Since (A2) implies the transitivity of epistemic indifference, this example shows that (A2) is false.

The above argument also shows that epistemic preferability cannot be given a probabilistic interpretation. If 'pPq' is read as 'p is more probable than q' then the standard definition of indifference as $\sim(pPq) \,\&\, \sim(qPp)$ forces us to read 'pIq' as 'p and q are equiprobable'. It would then follow that the epistemic indifference relation is transitive. Since this relation has already been shown to be nontransitive, the probabilistic interpretation of epistemic preferability must be rejected.

*Michigan State University,*
*East Lansing, Michigan 48823, U.S.A.*

© ROY A. SORENSEN 1983

# SMITH'S DEFENCE OF LEWY

*By* A. J. DALE

IN a recent paper Peter Smith has attempted to defend Lewy's contention that the Anderson–Belnap approach to entailment cannot deal with the paradoxes which Lewy discovered and published in his book *Meaning and Modality* (Peter Smith, 'In Defence of Lewy', ANALYSIS 42.3, June 1982, pp. 119–24). The contention itself had been attacked by John Slaney since in each paradox Lewy appeals to a principle of suppression of necessary premises (John Slaney, 'Entailment Survives the Lewy Paradoxes', ANALYSIS 41.4, October 1981, pp. 188–91). In his account Lewy distinguishes between a general application of the principle which would allow the derivation of $P \to Q \vee {\sim}Q$ from $P \cdot (Q \vee {\sim}Q) \to (Q \vee {\sim}Q)$ and a principle restricted in its application to those contexts where the premises (other than the necessary premise) and the conclusion are alike contingent. In such a context Lewy claims that the principle does seem intuitively acceptable (Lewy op. cit., p. 119).

Slaney shows that this 'intuitively acceptable' principle leads to unacceptable conclusions *within* propositional logic and thus cannot support Lewy's own paradoxes of entailment. The argument runs, with P contingent and Q necessary

1 $P \cdot Q \to P \cdot Q$

2 $P \to P \cdot Q$ (P and $P \cdot Q$ are both contingent)

3 $P \cdot Q \to Q$

4 $P \to Q$ (from 2 and 3 by transitivity).

Thus any contingent entails any necessary proposition (Slaney op. cit., p. 189).

Now Smith claims that a more restricted principle can be formulated which cannot be undermined by such an argument. First he defines a *thoroughly* contingent proposition as a contingent proposition that is not a truth function with some non-contingent constituent; then follows the restricted principle:

> If $A \cdot B$ entails C, where A and C are thoroughly contingent and B is necessary, then A entails C (Smith op. cit., p. 121).

The choice of the word 'constituent' is not a happy one since it needs elucidation itself: I shall assume Smith intends a well-formed sub-formula.

But this definition does nothing to rescue the principle since it allows through Lewy's own formulation of the second Lewis paradox (Lewy op. cit., p. 110)! For consider the argument with thoroughly contingent P and Q:

a $P\cdot(Q\vee\sim Q)\rightarrow(P\cdot Q)\vee(P\cdot\sim Q)$

b $P\rightarrow(P\cdot Q)\vee(P\cdot\sim Q)$ (both P and $(P\cdot Q)\vee(P\cdot\sim Q)$ are thoroughly contingent)

c $(P\cdot Q)\vee(P\cdot\sim Q)\rightarrow(Q\vee\sim Q)$

d $P\rightarrow(Q\vee\sim Q)$ (from b and c by transitivity)

So the principle revised in Smith's way is of no help to Lewy.

It may be that Smith intended some other interpretation of the phrase 'non-contingent constituent' but it should be remembered that Lewy's own definition of 'purely contingent' proved to be vacuous and that other theories of entailment designed to correct that definition have proved to have more difficulties attendant upon them (Michael Clark, 'Lewy's Doubts about Tantological Entailment', ANALYSIS 39.1, January 1979, pp. 30–4; A. J. Dale, 'Entailment and Substitution Instances', ANALYSIS 40.1, January 1980, pp. 10–12).

*The University,*
*Hull HU6 7RX*

© A. J. DALE 1983

---

# CAMBRIDGE PROPERTIES AND DEFUNCT INDIVIDUALS: A REPLY TO C. J. F. WILLIAMS

*By* BARRY MILLER

COMMENTING on my 'Negative Existential Propositions', C. J. F. Williams has raised two objections, one defending his denial of a distinction between internal and external negation, the other insisting that not all Cambridge properties can be acquired by an individual after it has ceased to exist. My reply, at its briefest, is that even if his points were correct, neither of them would make the slightest difference to my conclusion.

Foreseeing that some might continue to deny a distinction between internal and external negation, it was not by accident that I chose to discuss not only my own view that the negation in 'Arkle does not exist' was external, but also Williams' view that it was not. The results, however, were the same: in neither case were we committed to 'Arkle's acquiring any property that is real' (p. 188). So far as my conclusion is concerned, therefore, it is irrelevant whether there is or is not any distinction between internal and external negation. Thus, while Williams' point is of general interest, it has no particular bearing on my conclusion.

The same comment cannot be made about his second objection, for, since I did in fact say that 'there is nothing whatever paradoxical about a deceased Arkle acquiring a Cambridge property', he was entitled to think that a counter-example would be enough to invalidate my conclusion. Hence his suggestion that, although being shorter than my brother is a Cambridge property, it is not one that I can acquire after I cease to exist. About this I make two points, the first being that I am not inclined to agree that being shorter than my brother is a property that I could not acquire after ceasing to exist. Consider the case of two brothers, Tom and Bill, the latter having ceased to exist several years ago, in which time Tom has grown many centimetres. There would seem to be nothing paradoxical in saying, 'Tom has grown to be taller than Bill', nor in saying, 'It has come to pass that Tom is taller than Bill'. But, if the latter is lacking in paradox, then so is its alternative rendering, viz. 'It has come to pass that Bill is shorter than Tom'. And, if that is correct, there is no paradox in the property of being shorter than Bill being acquired by Tom after his ceasing to exist.

My second point is that, so far as the conclusion of my paper is concerned, it would make no difference even if I were wrong, and Williams' counter-example were correct. For, while I was careless enough to make a stronger claim than necessary (viz. that *all* Cambridge properties can be acquired without paradox by individuals that have ceased to exist), my purposes would have been served equally well by the more modest claim that *some* properties are such that they can be so acquired. The property of being famous is an obvious example. The property of non-existence is even more obvious, for although being famous can be acquired either before or after an individual's decease, it would be self-contradictory to suggest that non-existence could be acquired other than after an individual's ceasing to exist. Should Williams wish to deny my more modest, but sufficient, claim that some Cambridge properties are such that they can be acquired by a defunct individual, it would be pointless to offer any counter-examples; he would have to show that *no* Cambridge property can be acquired after an individual's decease.

Just as, for present purposes, it mattered not whether or not there was a distinction between internal and external negation, so, in the case of the property of being shorter than my brother, it matters not whether that particular property can or cannot be acquired after my decease. In neither case have we to allow that

'Arkle does not exist' generates the paradox that was thought to count against accepting '____ exists' as a first-level predicable.

*University of New England,*
*Armidale, N.S.W. 2351, Australia*

© BARRY MILLER 1983

# DISTANCE

*By* WILLIAM CHARLTON

A MATERIAL object has a place, or is somewhere, relatively to a reference-item, and we say where it is by giving its distance from that item in a specified direction. York, for instance, is 200 miles from London in the direction of the North Pole. My concern here is not with direction but only with distance: where *A* and *B* are material objects, what is it for *A* to be a mile from *B*?

According to Descartes it is for there actually to be a mile of matter between them: 'It is altogether repugnant that there should be any extension which is the extension of nothing' (*Principles* II.16). 'Yard', 'acre' and 'gallon' are words for units of length, area and volume, units of extension in one, two or three dimensions. If Descartes means to say that any actual yard, acre or gallon must be a yard of string, an acre of grass, a gallon of water, or the like, I am inclined to agree; but must there be an actual mile of this kind between things which are a mile apart?

Recent philosophical writers on space say little specifically about distance.[1] If they consider how we measure distance, they do not ask what a measurement of distance is a measurement of. The general assumption, however, seems to be that the only alternative to explaining distance in terms of actual lengths of material (or of absolute space) is to explain it in terms of possible lengthy objects.[2] In *Space and Time*, Ch. 4, Richard Swinburne says:

> We define a certain specified rigid rod . . . to be one unit of distance . . . The distance from *A* to *B* is then given by the smallest number of times we need to place such a unit rod along a line from *A* to *B* in order to reach *B*.

---

[1] Quine's account of length in *Word and Object* s. 50 could be adapted to distance as follows: 'Boston is 200 miles from New York' becomes 'For some *x*, *y* and *z*, *x* is 200, *y* is Boston, *z* is New York, and *x* is the distance in miles between *y* and *z*'; but we should then want to hear more about the triadic relationship of being the distance in miles between.

[2] I do not find any other alternative in, for instance, I. Hinkfuss's *Existence of Space and Time* or G. Nerlich's *The Shape of Space*. The use of the notion of continuous rigid movement made by J. Earman in 'Kant and the Nature of Space' (*Ratio* 1971) and L. Sklar in 'Substantiviality of Space' in *Journal of Philosophy* 1974 is completely different from the use of the notion of movement I defend here.

Certainly if 1,760 yard-long rods will just fit between *A* and *B*, it follows that *A* is a mile from *B*, and conversely. But does '*A* is a mile from *B*' just mean 'There could be a mile-long body between *A* and *B*' or 'There is room, space, for a mile of material between *A* and *B*'?

In my article 'Time' (*Philosophy* 1981) I point out that we think and speak of movements as measurable both in distance-units and in time-units, and I argue that (i) conceived as a mile long a movement is conceived as something which can be made many times, whereas conceived as an hour long a movement is conceived as a particular making or going on of a movement of so many feet or miles; (ii) the notion of a time-stretch is the notion of a going on of a movement or other sort of change. If those conclusions are correct, it is tempting to suggest that the notion of a distance (which is pretty much the notion of a stretch of space) is that of a movement which is or can be made. Considered as something which is or can be made, a movement is a possibility which is or can be fulfilled: it is a possible movement. The suggestion, then, is that the notion of a distance between bodies is the notion of a possible movement. '*A* is a mile from *B*' means 'The movement from *A* to *B* is a mile long'. Taking this view we can explain the notion of a space for a body in terms of movement: there is room for a yard-long body between *A* and *B* if the movement from *A* to *B* is a one-yard movement. I shall not here maintain the general thesis that distance (or space) should be explained in terms of possibilities rather than actualities; but I shall argue that this movement view, as I shall call it, has advantages over the length view, the view that the distance between *A* and *B* is the length of the body that would just bridge the gap.

First, there is a difference between distance and length which the length view obscures and the movement view illuminates. Because we use the same words 'yard', 'metre', etc., for units of distance and for units of length, and because if a pole is six feet long the distance from one end to the other is six feet, this difference is easily overlooked. It appears, however, in language. I do not mean simply that 'distance' is a kind of relational term, distance being distance *from* something, and 'length' is not. I mean that we speak of distances as moved and lengths as moved over or through. Your horse gallops a mile or the one-mile distance; it does not gallop *over* a mile. But it gallops *over*, and does not *gallop*, a mile of turf.

It may be objected that we say 'He swam the length of the pool'. Either, however, we mean he swam from end to end, and not from side to side; or we use 'the length of the pool' to express distance, as in 'A child who cannot swim the length of the pool is not safe in a canoe'. In talk about swimming 'length' is often used as a word for a unit of distance: 'I swim six lengths before breakfast'.

To swim a river is to swim across it, and rivers exist independently of our swimming them. They are external objects, so to speak, of swimming. Yards of carpet exist independently of our walking over them. But movements are not in the same way actual, they do not occur,[3] independently of being made. They are internal objects of making. How is it with distances? Do they exist untravelled? Only, I suggest, in the way in which unmade movements exist or are there to be made. For suppose I run a mile: is the mile I run more like the journey I make or like the race-track? Its status is that of a one-mile race. For to run a race is precisely to run a distance in competition with others. A one-mile race is either a one-mile distance run competitively, or a competitive running of a mile. But a race is an internal object of running. A unrun race may exist as a kind of possibility, but not as something actual. Similarly, then, with distances: the miles you move are the internal object of your moving.

Units of length can be defined by paradigmatic material objects. The length a yard is the length exemplified at a certain time by a certain bar.[4] If we take the movement view of distance, we must give a different definition of the distance a yard. It will have to be defined by a movement, and since bodies from which and to which there can be movement can themselves be displaced, it is most satisfactory to define distance-units, like time-units, by movements caused in a certain way. It would be possible to define the distance a mile as one 250,000th of the distance you have to move to get from here to the Moon, but it is less parochial to define it as one 186,000th of the distance light moves in a second. If we do that, of course, we must find some way of defining a second otherwise than as the time it takes light to move 186,000 miles, but that is not too hard.

A second advantage of the movement view is that it gives distances the status we feel they ought to have. If *A* is a mile from *B* the mile between them should be something real, yet not a third thing additional to them. Now the movement which starts at *A* and stops at *B* is a real possibility if it is really possible to move from *A* to *B*. The journey from London to York is a real journey in this way, whereas the journey from Vienna to Streslau in Ruritania is not real but fictitious. The journey from London to York is also really 200 miles long in that it is, in truth, necessary to move 200 miles in a northerly direction to get from London to York. On the other hand the journey is not a third thing coordinate with the two towns. Even those who wish to quantify over movements and changes put them in a different logical category from material objects. It may be suggested that the Newtonian conception of space as something just like glass or air except that it cannot be affected or affect other things is a kind of category mistake. The Newtonian wants there to be something real between objects which

---

[3] I take it that occurring is for movements and changes what existing is for material objects; the point is developed in my 'Causation and Change' forthcoming in *Philosophy*.

[4] So C. Strang, 'Plato and the Third Man', *Aristotelian Society* Suppl. Vol. xxxvii (1963).

3589

are not in contact, but the only kind of reality he can think of is that which attaches to material objects. The length view does not direct his attention to any other kind of reality, but the movement view exploits the fact that changes are real as well as things.

It may be urged in defence of the length view that it too allows there really to be a mile between $A$ and $B$ without there being any third thing between them. There really is a mile between $A$ and $B$ not if there *is* a third, mile-long thing between them, but if there *could* be. I reply that the notion of possibility does not accommodate itself so conveniently to lengths as to movements. A movement which is or can be made is something possible in the fundamental sense that it is that the going on of which is actual motion – it is what is made or goes on – though a change which is possible in this way is also physically possible, or, for that matter, inevitable or impossible, for particular bodies in particular circumstances. A movement is measurable in distance-units just in so far as it is possible. Hence we can legitimately speak of miles of possible movement or say that between us and the Moon there are 250,000 possible miles of movement. Whilst, however, we might say that between us and the Moon there could be 250,000 miles of string, we cannot say that there are 250,000 possible miles of string or speak of miles of possible string. In general, miles of string or tarmac are actual. It might be said that a nautical mile and a roman mile are two possible miles, but there are not 250,000 possible miles of that sort between us and the Moon, and a possible 250,000-mile-long object is Quine's possible fat man in the doorway writ large. All we can say is that it is possible that there should be a 250,000-mile-long object between us and the Moon, and in what sense of 'possible' is this possible? The only answer seems to be that there is space or room for such a lengthy object; and the notion of space or room is primitive and unanalysable.

A final point. Questions like 'Is space infinite?' 'Is it Euclidean?' 'Is the distance between two bodies relative to an observer's frame of reference?' though they belong more to physics than to philosophy, attract philosophical interest. They sound less strange when construed as questions about possible movement than they would if taken to be about hypothetical lengthy objects. 'Is space infinite?' means rather 'Is there any limit to the number of miles it is possible to move?' than 'Is there any limit to the number of miles long a body for which there is room can be?' If we want to know whether space is Euclidean, we want to know if it is possible to move in a straight line; or, if we take a straight line to be a line in which a body moves when moving uniformly, whether two bodies moving uniformly can move parallel, without converging or diverging. The thesis that distance is relative is sometimes stated as the thesis that a moving body contracts in the direction in which it is moving. That makes it sound as if a body can be affected and change in size without being acted upon. We avoid the suggestion if we say that a movement from one end of a body to the other is shorter for, takes less time by the clocks of, an observer relatively to whom it is moving.

*The University,*
*Newcastle upon Tyne*
*NE1 7RU*

© William Charlton 1983

# TRANSLATION AND BELIEF AGAIN

*By* Arnold Cusmariu

IN an earlier paper ('Translation and Belief', Analysis 42.1, January 1982, pp. 12-16) I presented a two-stage version of Church's translation argument against Carnap's linguistic analysis of belief. I believe that the assumptions embodied in those two stages are necessary if the intent is to refute Carnap's analysis outright. However, it turns out that the first stage by itself is also sufficient to establish a weaker though no less significant conclusion provided a principle to the effect that the same thought or idea can be expressed, in some sense, in two different languages is added. The conclusion is this: for any two languages L and L′ between which there exists a tradition of correct translation, if Carnap's analysis of belief is correct as expressed in L, then it is incorrect as expressed in L′. This is not the conclusion Church sought but nonetheless it seems to be worthwhile to have another serious objection at hand based on the translation device which does not make use of the (possibly) controversial epistemic principle Q (see p. 14) I used at the second stage.

Consider the following imaginary but instructive case. Two philosophers, one English and one German, arrive at Carnap's analysis of belief independently of each other. Each formulates the analysis in his own language and neither knows what the other has accomplished. The English philosopher claims that E

E. There is a sentence Z in a semantical system S′ such that (a) Z in S′ is intensionally isomorphic to 'Snow is white' in English, and (b) Smith believes Z as a sentence of S′

is the correct analysis of A

A. Smith believes that snow is white,

and the German philosopher claims that E3

E3. Es gibt einen Satz auf einem semantischen System S′ so dass (a) Z als Satz von S′ intensional isomorph zu 'Der

> Schnee ist weiss' als Deutcher Satz ist, und (b) Smith Z als Satz von S' glaubt

is the correct analysis of A'

> A'. Smith glaubt, dass der Schnee weiss ist.

One day they meet at a symposium and are delighted to discover that they had stumbled upon the same idea. Just to amuse himself, the German philosopher translates A and E into German and, being careful to do so correctly, obtains A' and E'

> E'. Es gibt einen Satz auf einem semantischen System S' so dass (a) Z als Satz von S' intentional isomorph zu 'Snow is white' als Englischer Satz ist, und (b) Smith Z als Satz von S' glaubt.

This done he returns to his own work, E3, reflects for a moment and becomes greatly puzzled. His puzzlement is such that he is momentarily reluctant to share it with his English colleague. The question is, why?

The answer depends on a rather cumbersome principle which I will first state in the abstract, postponing discussion of it until after I have applied it to our case:

> Q*. If X is a correct analysis of Y as expressed in language L, and if X' is a correct analysis of Y' as expressed in language L', and if X and X' 'express the same idea', and if X" and Y" are correct translations of X' and Y' into L, then, X should contain the same information as X" and Y should contain the same information as Y".

To fill this in with the details of our scenario (and here I beg the reader's indulgence), let X = E3, Y = A', L = German, X' = E, Y' = A, L' = English, X" = E' and Y" = A'. Now, I suggest that our man became puzzled and was at first reluctant to share his puzzlement with his English colleague because (i) he first realized that E3 and E' do not contain the same information, by P*

> P*. Sentence X of natural language L contains the same information in L that sentence X' of natural language L' contains in L' only if X implies whatever X' implies,

since E3 implies that Z is intentionally isomorphic with a German sentence but E' does not; he then assumed (ii) that he had translated correctly from English into German, (iii) that his own analysis as expressed initially in German was correct and (iv) that he and his English colleague had achieved the same result. These reflections together with Q* then led him to the paradoxical conclusion that his colleague's analysis as expressed in English, E, had to be incorrect! Of course, had he revealed this state of affairs to his colleague the latter would immediately have reasoned the same way, only to a very different and embarrassing conclusion; hence the momentary reluctance to share his puzzlement.

In Q* it is asserted, what seems plausible, that sameness of information is a necessary condition of correct intertranslation between correct analyses expressing the same thought or idea. If this idea that the same philosophical theory can be expressed in two different languages seems controversial in the present context, I can take some of the sting out by noting that no particular explanation of it need be presupposed here. Anyhow, as I do not see that Carnap's theory of belief entails any particular account of the nature of philosophical theories, there is no real danger of begging the question against him in Q* however the idea is analysed.

*Texas A &M University,*
*College Station, Texas 77843, U.S.A.*

© ARNOLD CUSMARIU 1983

---

# DRETSKE ON KNOWLEDGE AND INFORMATION

*By* J. CHRISTOPHER MALONEY

FRED DRETSKE, in his recent and sure to be influential *Knowledge and the Flow of Information* (Cambridge, Mass., 1981; designated *KFI*), articulates a new and promising theory of knowledge, belief and conceptualizaton grounded in information theory. Dretske's ideas should share in shaping developments in both epistemology and the philosophy of mind. Nevertheless, I shall argue that Dretske's analysis of knowledge is subject to an objection bearing on one of the distinctive features of his account which differentiates it from traditional treatments of knowledge as justified true belief.

Restricting his proposal to *de re* perceptual knowledge of contingent facts, Dretske maintains that such knowledge is belief caused by information. More precisely and with the restrictions suppressed,

(1) *K* knows that *s* is *F* iff *K*'s belief that *s* is *F* is caused (or causally sustained) by the information that *s* is *F* (*KFI*, p. 86).

We need not here attend to the concept of a belief's being causally sustained. '*s*' is a demonstrative in a *de re* context characteristic of that kind of perceptual knowledge of *s*'s being *F* independent of what else *K* might believe or fail to believe of *s*. Nothing, then, depends on considerations regarding *s*'s identity.

But everything does depend on the notion of a particular piece of information causing a specific belief. Dretske distinguishes between the information that *s* is *F* and a physical signal *r*, an event, state or condition, which, in virtue of having some property *F**, conveys or carries the information that *s* is *F* (*KFI*, p. 87). We can imagine, modifying one of Dretske's examples, that *r* is a knock on the door which, in virtue of being a pair of three raps ($F^*$), carries the information that this (*s*) is the courier (*F*).

Since information is abstract, its causal efficacy is, according to Dretske, traced to the causality of its signal. That is,

> (2) When a signal carries the information that *s* is *F* in virtue of having the property $F^*$, then and only then the information that *s* is *F* causes whatever the signal's being $F^*$ causes (*KFI*, p. 87).

We should pause to notice a consequence of (2). During the bottom of the ninth inning in a tied baseball game, a particular fan's jumping to his feet and wildly cheering carries, as a signal, the information that his team has just won. This same behaviour also happens to startle his disinterested companion into spilling his beer. (2) entails that the information that the fan's team has won caused his companion to spill his beer. Yet this seems incorrect. The overt behaviour, not the information it conveys, causes the mishap, especially if the companion remains ignorant of the course of the game.

Dretske might reply that the property of the fan's behaviour that carries the information that the team has won is not the property of the behaviour that causes the companion to spill his beer. Perhaps, though that would appear *ad hoc*. Nevertheless, the point of the example is clear. One and the same property of an event can launch different causal chains, and this, together with (2), may force us to attribute spurious effects to information. Accordingly we might endorse an amended version of (2):

> (2′) When a signal, *r*, carries the information that *s* is *F* to *K* in virtue of *r*'s being $F^*$, then and only then the information that *s* is *F* causes whatever beliefs of *K* *r*'s being $F^*$ causes.

No matter which of (2) or (2′) we entertain, it remains to acknowledge Dretske's explanation of the conditions under which a signal carries information. He proposes that

> (3) A signal, *r*, carries the information that *s* is *F* (to *K*) iff the conditional probability of *s*'s being *F*, given *r* (and what *K* already knows about the possibility of *s* being *F*), is 1 (but, given only what *K* knows independent of receiving *r*, is less than 1) (*KFI*, p. 65). (Parenthetical reference to *K* is mine.)

Let us note that Dretske argues persuasively that the probability, of information must be 1, relative to a signal, for the signal to convey the information (*KFI*, pp. 99–106). Conveyed information must thus be true (*KFI*, pp. 43–45). (3) together with either (2) or (2′) explicate the concepts essential to (1). Dretske also defuses concern that reference to *K*'s knowledge in (3) renders (1) circular, citing the recursive character of (1) (*KFI*, p. 86).

With (1)-(3) in hand, we can now consider a serious objection to Dretske's information-theoretic analysis of knowledge. Suppose that *K* lives in a relatively advanced scientific culture, isolated from but otherwise much like our own, differing from our culture only in selected scientific discoveries. Within *K*'s culture theoretical physics, for example, is better understood than in our own, but water, though plentiful, has received little attention from the scientists of *K*'s culture. As a result, they do not know that water is $H_2O$. *K* is not a scientist but is generally well informed. He avidly reads science fiction which, just as we know it, contains some scientifically demonstrated truths as well as pure speculation. But not all of this speculation need be false. In a particular work some of the speculation might happen to be true though, of course, not known to be so. *K*'s favourite piece of science fiction offers the idea that water is $H_2O$. *K*, knowing that the science of his society has been silent on the matter, comes to believe on the basis of his reading that water is $H_2O$. *K*, then, is rather like us, in that the literature he reads affects his beliefs. Here we assume that *K*'s reading the book in question causes him to believe that water is $H_2O$. And importantly, as it happens, all that the book speculates about water, relevant to *K*'s belief, is true. Thus, *K*'s belief about water not only is true but also depends on no false beliefs. (Compare *KFI*, p. 98.)

One day shortly after coming to believe that water is $H_2O$, *K*, while walking in a drizzle, sees and feels a raindrop strike his extended palm. His sensing this raindrop is, according to Dretske, a signal which, by (3), carries the information that this (raindrop) is water. For, given the normality of *K*'s present situation, his sensory signal (and what else he and everyone else know about rain and water), the probability that this (particular raindrop) is water is 1 (*KFI*, pp. 123-34). Although unknown in *K*'s culture, it remains a genuine law of nature that water is $H_2O$. Hence, since the probability that this (raindrop) is water is 1, given *K*'s sensory signal, the probability that this (same raindrop) is $H_2O$ is also 1, relative to the same sensory signal. Thus, and again by (3), *K*'s sensory signal also carries the information that this (still the same raindrop) is $H_2O$ (*KFI*, p. 73). Crucially, believing as he does that water is $H_2O$, *K*'s signal causes him to believe of this very raindrop that it is $H_2O$. So, by either (2) or (2′), the information about the raindrop that it is $H_2O$ causes *K* to believe of the raindrop that it is $H_2O$. Therefore, by (1), *K knows* of the raindrop that it is $H_2O$. Surely, however, this is false. Thus, at least one of (1)-(3) must be incorrect. Yet collectively they constitute the core of Dretske's information-theoretic analysis of knowledge.

*K* does not know of the raindrop that it is $H_2O$, because his so believing lacks the appropriate epistemic justification, despite depending on only true beliefs. The absence in Dretske's theory of the vexing and not yet adequately understood notion of justification is, I think, problematic. Dretske, however, considers it a mark of distinction that his theory does not depend upon an unexplicated concept of epistemic justification (*KFI*, p. 85), and it would indeed be a boon to epistemology should he be right. Perhaps Dretske can convincingly argue that his theory is immune to the present problem by appealing to either the epistemic status of laws of nature (*KFI*, p. 86, n. 2) or the character of the concepts occurrent in *K*'s cognitive processes (*KFI*, pp. 215-18). But since, *ex hypothesi*, *K*'s relevant background beliefs are all true, nothing in the example trades on the epistemic effects of false beliefs (*KFI*, pp. 96-99). In any event, from Dretske we can, as always, expect clear signals full of information.

*Oakland University,*
*Rochester, Michigan 48063,*
*U.S.A.*

© J. CHRISTOPHER MALONEY 1983

---

# TRACKING NOZICK'S SCEPTIC: A BETTER METHOD

*By* RAYMOND MARTIN

EPISTEMOLOGISTS have generally thought that justified belief is necessary for knowledge. However, Robert Nozick has recently defended an analysis of knowledge according to which justified belief appears not to be necessary for knowledge.[1] Nozick expresses some hesitation about whether his analysis does have this implication. I wish to argue that it does, and also that, probably because it does, it is an unacceptable analysis of knowledge. Nozick uses his analysis of knowledge as a basis for an ingenious response to scepticism. The difficulties with his analysis cast a shadow on the acceptability of that response.

On Nozick's analysis, S knows that *p* if and only if

(1) *p* is true.
(2) S believes, via method or way of coming to believe M, that *p*.
(3) If *p* weren't true and S were to use M to arrive at a belief whether (or not) *p*, then S wouldn't believe, via M, that *p*.
(4) If *p* were true and S were to use M to arrive at a belief whether (or not) *p*, then S would believe, via M, that *p*. (p. 179)

---

[1] Robert Nozick, *Philosophical Explanations* (Cambridge, Massachisetts, 1981). All page numbers noted parenthetically in the present paper, except the last page number noted, are to this book.

Nozick's accounts of (3) and (4) are sketchy. Simplifying somewhat, Nozick says that (3) speaks of those not-*p* worlds in which S uses M to arrive at a belief whether (or not) *p* that are 'closest to' (or 'in the neighborhood of') the actual world. (3) says of those worlds that, in them, S does not believe, via M, that *p*. Similarly, (4) speaks of those *p* worlds in which S uses M to arrive at a belief whether (or not) *p* that are 'in the neighborhood of' the actual world. (4) says of those worlds that, in them, S believes, via M, that *p*. No definite account is given of the notion of 'neighborhood' (but see pp. 680–1, n. 8). Nozick admits that his explanations of (3) and (4) may not be helpful, and suggests that the reader follow his own intuitions (p. 176).

Nozick says of someone who believes that *p*, when *p* is true, and when (3) and (4) also obtain, that his belief that *p* *tracks* the truth that *p*: 'Knowledge is a particular way of being connected to the world, having a specific real factual connection to the world: tracking it.' (p. 178). Nozick uses his analysis of knowledge and the related notion of 'tracking' as the basis for his response to scepticism.

The central thrust of Nozick's response to scepticism may be conveyed by considering how Nozick would respond to the following traditional sceptical argument:

(1) S knows that if he is now seated at the desk in his study, then he is not now in bed hallucinating that he is now seated at the desk in his study.
(2) S does not know that he is not now in bed hallucinating that he is now seated at the desk in his study.
Hence, (3) S does not know that he is now seated at the desk in his study.

Since Descartes most philosophers who have rejected this sort of sceptical argument have argued against (2). But repeated attempts to explain why (2) is false have failed to convince each new generation of epistemologists. Nozick's strategy is different. He concedes that (2) is true, but denies that the argument is valid.

The sceptical argument presupposes that knowledge is closed under known logical implication, that is, that if a person knows some proposition *p*, and also knows that *p* logically entails *q*, then he also knows that *q*. Nozick denies that knowledge is closed under known logical implication. But if it isn't, what explains how it is possible both that S knows that he is now seated at the desk in his study, yet does not know that he is not now in his bed hallucinating? Nozick's answer is his analysis of knowledge: S's belief that he is now seated at the desk in his study tracks the fact that he is now seated at the desk in his study, even though it does not track the fact that he is not now in his bed hallucinating that he is now seated at the desk in his study. This seems to work, provided that Nozick's analysis of knowledge is acceptable.

But is Nozick's analysis of knowledge acceptable? A problem for it is that it seems that someone could have a belief that *p* which tracked the fact that *p* even though he was not justified in believing that *p*. This might happen in case he came to believe that *p* via an unreliable method, one which was unlikely to produce mostly true beliefs. Nozick is uncertain what to say about this possibility, but leans toward suggesting that if someone arrives at his belief that *p* via an unreliable method, yet his belief that *p* tracks the truth that *p*, then his unreliable method must somehow involve a submethod that is reliable.

> . . . whenever someone knows that *p*, will he also have a justified true belief that *p*? Not if a person can believe via a method that is highly unreliable (and less reliable than others available to him) yet which not only yields the truth that time but also tracks that one time. However, since tracking involves some generality to other (subjunctive) situations, it seems plausible that whenever a person knows, there will be some reliable submethod via which he knows. So it seems plausible that justified true belief is a necessary condition for knowledge; but I prefer to leave this question unsettled. (p. 267)

But it does not seem that the sort of generality involved in tracking insures that the method one uses to arrive at a belief which tracks the truth either is reliable or in some way involves a reliable submethod.

Consider the following example:[2] Suppose that S bets on a horse at his local track. The bet S places is such that if S picks the winner in the first race or the winner in the second race, or both, S wins ten dollars; otherwise, S loses. S picks Gumshoe in the first and Tagalong in the second. But S has the following peculiar method for arriving at his belief whether (or not) Gumshoe wins the first: S shields himself from any information about the results of either of the first two races, and then, after the conclusion of the second race, S presents his ticket to an automated cashier. (The cashier either pays S for his ticket or not, but cannot, due to the way it is constructed, provide S with any additional information about the results of either race.) If the cashier pays S for his ticket, then S believes that Gumshoe won the first; otherwise, S does not believe that Gumshoe won the first. On the occasion under discussion, Gumshoe wins the first and Tagalong finishes last in the second. Tagalong's performance in the second, moreover, was causally independent of Gumshoe's performance in the first: no matter how Gumshoe had performed in the first, Tagalong still would have finished last in the second. After the second race, with no information on the results of either race, S goes to the cashier and presents his ticket. The cashier gives S ten dollars in exchange for S's ticket. The cashier would not have given

---

[2] A similar example poses a problem for Fred Dretske's similar response to scepticism. See my, 'Empirically Conclusive Reasons and Scepticism', *Philosophical Studies*, 28 (1975) 215–17.

S ten dollars in exchange for S's ticket unless *either* Gumshoe *or* Tagalong had won. But the cashier conveys no information to S about whether it was Gumshoe or Tagalong, or both, that won.

S, in our example, satisfies all of the requirements of Nozick's analysis: Gumshoe won the first; S believes, via M, that Gumshoe won the first; if it weren't true that Gumshoe won the first and S were to use M to arrive at a belief whether (or not) Gumshoe won the first, then (since the cashier would not then have paid S for his ticket) S wouldn't believe, via M, that Gumshoe won the first; and if it were true that Gumshoe won the first and S were to use M to arrive at a belief whether (or not) Gumshoe won the first, then (since the cashier then would have paid S for his ticket), S would believe, via M, that Gumshoe won the first. But S arrived at his belief that Gumshoe won the first via an unreliable method and one which, it seems, does not involve any submethod which is reliable. Moreover, it is seriously counterintuitive to suppose that S *knows* that Gumshoe won the first. Thus, Nozick's analysis of knowledge has a seriously counterintuitive consequence.

There may be some uncertainty about whether S, in our example, satisfies Nozick's condition (4). Nozick says that

> Condition (4) is an adherence condition, so the relevant doubts concern how securely you are tied to the truth. For many (most?) of the things *p* you believe, if a group of people come and deceitfully told you not-*p*, you would believe them and stop believing *p*. . . . So do you really know *p*? (p. 212)

Nozick's answer is that you still might really know *p*. The fact that if certain things happened, you would not believe *p* is relevant to whether you know *p* only if those things happen in 'the first part of the *p* neighborhood of the actual world' (p. 212). But they *might not* happen in 'the first part of the *p* neighbourhood of the actual world'. As Nozick puts it, 'it is fortunate for my knowing that *p* that there wouldn't be people who trick me . . .' (p. 212). Although I find Nozick's explanation of (4) obscure, I can find in it no reason why S, in our example, should not be 'securely tied to the truth' that Gumshoe won the first. Or, if S, in our example, might not satisfy condition (4), there seems to be no reason why we could not elaborate our example so as to insure that S does satisfy condition (4).

Nozick might be tempted to reply (compare p. 186) that S, in our example, does not satisfy (4) because were S to use the same method to arrive at a belief whether (or not) Gumshoe won the first in a possible world in which Gumshoe won the first but the automated cashier for some reason – say, because it malfunctioned – did not pay S for his ticket, S would not believe that Gumshoe won the first. But such an interpretation of (4) is too stringent. (4) then would fail to be satisfied by many examples in which a person really does arrive at knowledge through the use of a reliable instrument.

Our example seems to show that satisfying the four requirements of Nozick's analysis does not insure either that S knows that *p* or that S arrived at his belief that *p* in any way that involves S's use of a reliable method. Suppose that it does show these things. How then to patch up Nozick's analysis? The obvious temptation is to modify the analysis so as to require that S arrives at his belief that *p* via a reliable method. But how should this modification be written into Nozick's analysis? And what exactly is a reliable method? And will Nozick's analysis, so modified, provide as suitable a vehicle as his original analysis for his response to scepticism?

There are no simple answers to these questions. It is difficult to answer the last question until one knows exactly how Nozick's analysis should be modified. And if it should be modified by requiring that S came to believe that *p* via a reliable method (or alternatively, via the most appropriate reliable method), which seems the only hopeful modification, then it is difficult to answer the last question above until one can specify what is meant by 'a reliable method'. But there are formidable problems, as Nozick himself acknowledges (pp. 264-5), to specifying what is meant by 'a reliable method'. So, if Nozick's analysis needs to be modified by requiring that S came to believe that *p* via a reliable method, then it must remain an open question, at least until we can specify acceptably what is meant by 'a reliable method', whether Nozick's response to scepticism is acceptable.

It may seem that there are ways to modify Nozick's analysis in order to avoid our counterexample that do not introduce into his analysis the notion of 'a reliable method' or any related notion that is problematic for the same reasons. I have been unable to think of any such way which does not lead quickly to unacceptable consequences. But, of course, there may be such a way. Suppose, for the sake of argument, that there is no such way. Then a basic flaw in Nozick's analysis of knowledge seems to be that having a belief which tracks the truth does not insure that one is justified in having that belief, and for that very reason, does not insure that one has knowledge. If our supposition is correct, then justified belief does seem to be necessary for knowledge. However, the sort of justified belief which is thus shown to be necessary for knowledge is not necessarily the sort that has been stressed by epistemologists in the Cartesian tradition. According to accounts in that tradition, S knows that *p* only if S is justified in believing that *p*, and S is justified in believing that *p* only if either *p* is self-warranting for S or *p* is adequately supported by beliefs each of which is self-warranting for S.[3] Specifying acceptably what is meant by 'S's coming to believe that *p* via a reliable method' may make no reference to the notion of 'self-warranting beliefs'. This is important since the notion of 'self-warranting beliefs' has been notoriously troublesome and a source of much epistemological mischief.

---

[3] For a fuller treatment of the differences between Cartesian and reliability accounts of justification, see Alvin Goldman, 'Discrimination and Perceptual Knowledge', *The Journal of Philosophy*, 73 (1976) 771-91, and Richard Grandy, 'Ramsey, Reliability and Knowledge', in D. H. Mellor (ed.), *Prospects for Pragmatism* (Cambridge, 1980), pp. 197-210.

The idea that reference to reliable methods or processes plays a role in the analysis of knowledge and that it may replace reference to Cartesian-type justification, while not entirely new (the idea was apparently suggested by Ramsey in 1929),[4] seems to be an idea whose time has come. Richard Grandy, in a useful survey of some of the relevant literature, finds anticipations of the idea in work by Watling and by Unger, and a more developed form of it in work by Armstrong.[5] Grandy finds decisive objections to the analyses of each of these philosophers, but he argues persuasively that they are on the right track in stressing the epistemological importance of reliability over Cartesian-type justification. In the end, however, he confesses that 'we have thus far only a very poorly developed theory of reliability' (p. 209). That conclusion seems to be right. And until we develop a more adequate theory of reliability, epistemologists properly wary of promissory notes will remain sceptical whether the notion of reliability may be incorporated into an analysis of knowledge so as to provide the basis for a satisfactory response to scepticism.

*University of Maryland,*
*College Park, Maryland 20742, U.S.A.*

© RAYMOND MARTIN 1983

---

[4] F. P. Ramsey, *Foundations*, ed. by D. H. Mellor (London, 1978), pp. 126–7.

[5] Grandy, *op. cit.* Further developments of the idea may be found in Goldman, *op. cit.*, and in Fred Dretske, 'Conclusive Reasons', *The Australasian Journal of Philosophy*, 49 (1971) 1–22.

---

# THE TOXIN PUZZLE

### *By* GREGORY S. KAVKA

YOU are feeling extremely lucky. You have just been approached by an eccentric billionaire who has offered you the following deal. He places before you a vial of toxin that, if you drink it, will make you painfully ill for a day, but will not threaten your life or have any lasting effects. (Your spouse, a crack biochemist, confirms the properties of the toxin.) The billionaire will pay you one million dollars tomorrow morning if, at midnight tonight, you *intend* to drink the toxin tomorrow afternoon. He emphasizes that you need not drink the toxin to receive the money; in fact, the money will already be in your bank account hours before the time for drinking it arrives, if you succeed. (This is confirmed by your daughter, a lawyer, after she examines the legal and financial documents that the billionaire has signed.) All you have to do is sign the agreement and then intend at midnight tonight to drink the stuff tomorrow afternoon. You are perfectly free to change your mind after receiving the money and not drink the toxin. (The presence or absence of the intention is to be determined by the latest 'mind-reading' brain scanner and computing device designed by the great Doctor X. As a cognitive scientist, materialist, and faithful former student of Doctor X, you have no doubt that the machine will correctly detect the presence or absence of the relevant intention.)

Confronted with this offer, you gleefully sign the contract, thinking 'what an easy way to become a millionaire'. Not long afterwards, however, you begin to worry. You had been thinking that you could avoid drinking the toxin and just pocket the million. But you realize that if you are thinking in those terms when midnight rolls around, you will not be intending to drink the toxin tomorrow. So maybe you will actually have to drink the stuff to collect the money. It will not be pleasant, but it is sure worth a day of suffering to become a millionaire.

However, as occurs to you immediately, it cannot really be necessary to drink the toxin to pocket the money. That money will either be or not be in your bank account by 10 a.m. tomorrow, you will know then whether it is there or not, and your drinking or not drinking the toxin hours later cannot affect the completed financial transaction. So instead of planning to drink the toxin, you decide to intend today to drink it and then change your mind after midnight. But if that is your plan, then it is obvious that you do not intend to drink the toxin. (At most you intend to intend to drink it.) For having such an intention is incompatible with planning to change your mind tomorrow morning.

At this point, your son, a strategist for the Pentagon, makes a useful suggestion. Why not bind yourself to drink the stuff tomorrow, by today making irreversible arrangements that will give you sufficient independent incentive to drink it? You might promise someone who would not later release you from the promise that you will drink the toxin tomorrow afternoon. Or you could sign a legal agreement obligating you to donate all your financial assets (including the million if you win it) to your least favourite political party, if you do not drink it. You might even hire a hitman to kill you if you do not swallow the toxin. This would assure you of a day of misery, but also of becoming rich.

Unfortunately, your daughter the lawyer, who has read the contract carefully, points out that arrangement of such external incentives is ruled out, as are such alternative gimmicks as hiring a hypnotist to implant the intention, forgetting the main relevant facts of the situation, and so forth. (Promising *yourself* that you will drink the toxin could help if you were one of those strange people who take pride in never releasing oneself from a promise to oneself, no matter what the circumstances. Alas, you are not.)

Thrown back on your own resources, you desperately try to convince yourself that, despite the temporal sequence, drinking the toxin tomorrow afternoon is a necessary condition of pocketing the million that morning. Remembering Newcomb's Problem, you seek inductive evidence that this is so, hoping that previous recipients of the billionaire's offer won the million when and only when they drank the toxin. But, alas, your nephew, a private investigator, discovers that you are the first one to receive the offer (or that past winners drank less often than past losers). By now midnight is fast approaching and in a panic you try to summon up an act of will, gritting your teeth and muttering 'I will drink that toxin' over and over again.

We need not complete this tale of high hopes disappointed (or fulfilled) to make the point that there is a puzzle lurking here. You are asked to form a simple intention to perform an act that is well within your power. This is the kind of thing we all do many times every day. You are provided with an overwhelming incentive for doing so. Yet you cannot do so (or have extreme difficulty doing so) without resorting to exotic tricks involving hypnosis, hired killers, etc. Nor are your difficulties traceable to an uncontrollable fear of the negative consequences of the act in question – you would be perfectly willing to undergo the after-effects of the toxin to earn the million.

Two points underlie our puzzle. The first concerns the nature of intentions. If intentions were inner performances or self-directed commands, you would have no trouble earning your million. You would only need to keep your eye on the clock, and then perform or command to yourself at midnight. Similarly, if intentions were simply decisions, and decisions were volitions fully under the agent's control, there would be no problem. But intentions are better viewed as dispositions to act which are based on *reasons to act* – features of the act itself or its (possible) consequences that are valued by the agent. (Specifying the exact nature of the relationship between intentions and the reasons that they are based on is a difficult and worthy task, but one that need not detain us. For an account that is generally congenial to the views presented here, see Davidson's 'Intending', in his *Essays on Actions and Events*.) Thus, we can explain your difficulty in earning a fortune: you cannot intend to act as you have no reason to act, at least when you have substantial reasons not to act. And you have (or will have when the time comes) no reason to drink the toxin, and a very good reason not to, for it will make you quite sick for a day.

This brings us to our second point. While you have no reasons to drink the toxin, you have every reason (or at least a million reasons) to *intend* to drink it. Now when reasons for intending and reasons for acting diverge, as they do here, confusion often reigns. For we are inclined to evaluate the rationality of the intention both in terms of its consequences and in terms of the rationality of the intended action. As a result, when we have good reasons to intend but not to act, conflicting standards of evaluation come into play and something has to give way: either rational action, rational intention, or aspects of the agent's own rationality (e.g., his correct belief that drinking the toxin is not necessary for winning the million).

I made some similar points in an earlier article ('Some Paradoxes of Deterrence', *Journal of Philosophy*, June 1978), but there I was discussing an example involving conditional intentions. The toxin puzzle broadens the application of that discussion, by showing that its conclusions may apply to cases involving unconditional intentions as well. It also reveals that intentions are only partly volitional. One cannot intend whatever one wants to intend any more than one can believe whatever one wants to believe. As our beliefs are constrained by our evidence, so our intentions are constrained by our reasons for action.[1]

*University of California, Irvine,*
*California 92717, U.S.A.*

© GREGORY S. KAVKA 1983

---

[1] The puzzle discussed here emerged from a conversation, some years ago, with Tyler Burge about 'Some Paradoxes of Deterrence'. I have profitably discussed it with Paul Humphries, Rick O'Neil, and Virginia Warren, but am alone responsible for its present form and the conclusions derived from it. I am grateful to Doris Olin for suggesting a needed change in an earlier draft.

---

# JIM AND THE INDIANS

*By* MARTIN HOLLIS

I WOULD not be writing these memoirs but for a nasty incident in the small South American town of ——— in the summer of 19—. Indeed I would have kept quiet altogether, if a fellow called Bernard Williams hadn't got wind of it and spread a version which I cannot endorse.[1]

---

[1] The direct references are to Bernard Williams' part of J. J. Smart and B. Williams, *Utilitarianism: For and Against*, Cambridge University Press, 1973, esp. p. 98ff., and the oblique ones to 'Internal and External Reasons' in Williams' collection of essays *Moral Luck*, Cambridge University Press, 1981. I would like to thank Peter Hobbis for his helpful comments.

It was a sleepy small town, friends had told me, but, as I stumbled into the sunbaked square just before noon, I could see that something was afoot. The tense crowd and the rifles of a squad of sweating militia were the first sign. Then I spotted a trembling group of twenty Indians up against the post-office wall. I had interrupted an execution of some sort and the officer in charge was pretty put out. Captain Pedro they called him and he was not, let me tell you, a man to take on your honeymoon. Play was stopped and he set about questioning me. Why was I there? Did I admire the regime? Did I agree that the Indians were reaping great benefits from it? That this small lesson in loyalty would be to the advantage of all?

Even the first question was a facer. The truth was that I was there to sell vacuum cleaners, no one having told me what I had seen instantly on arrival, that there was no electricity. But that made me an ass, if he believed me, or a spy, if he didn't. Either way, the number of targets would be upped smartish to twenty one. So was I the country's millionth tourist? an international football star? an official of the Electoral Reform Society? In a flash of inspiration I claimed to be a botanist sent by the Royal Society to look for orchids at the Queen's personal request. His dark glasses gave nothing away but his moustache twitched, I fancied, respectfully. 'Indeed, Señor? I too hold degrees in botany. Let us sit down.' A snap of the fingers produced a table, chairs and a Coca Cola umbrella.

To my initial relief, his mind was not on botany. 'This is an historic occasion, Señor,' he began after some private calculation, 'and we must mark it suitably. So I invite you to shoot just one of these worthless Indians on Her Majesty's behalf and I, for my part, will then release the other nineteen as a gesture of regard and clemency. *D'accord?*'

My first idea was, of course, to knock him cold and free the Indians in a trice. But short and tubby vacuum cleaner salesmen know their limitations and it was my turn to think. Ethics never was my strong suit. I could dimly recall my grandmother saying, 'Do as you would be done by, Jim, and you won't go far wrong.' But that wasn't much help and, besides, I doubt if she meant it to apply to foreigners. On the other hand, since Pedro was hardly to be trusted, there wasn't much chance of saving any Indians by getting involved. My dismay must have shown, because he broke in on my thoughts. 'Even if, laughable to suggest, we did not trust one another, Señor, it would make no difference, since even a slight chance of saving nineteen lives is better than none. The Indians will certainly see it that way.' A pair of terrified Indians was beckoned over and quickly proved him right. 'You really have no choice. Like all true Englishmen, you no doubt believe in the greatest happiness of the greatest number. So the only question is how to estimate the remote consequences and we can safely say that, in any case, they are heavily outweighed by the immediate ones.'

'But my integrity . . .', I stammered.

'That gives no independent leverage at all, my dear chap. A good man works out the right thing to do and then his integrity demands simply that he goes ahead and does it. You don't *want* to shoot anyone, I daresay, but it would be self-indulgent to jib on that score.'

'But', I said, rallying a little, 'either *I* shoot one Indian or *you* shoot twenty. That's the point. I am morally responsible for what I do but not for what you cooked up in the first place and not for the twenty you shoot.'

'Come, come! If you were the only able-bodied person in a lifeboat with twenty helpless wounded and you knew that the boat would sink unless lightened by one, you would naturally tip one overboard. You wouldn't let the lot drown, because it was all no fault of yours. That is how things are here. I shall be responsible too, if that will make you feel better. But your presence makes it possible for us to save nineteen lives; so, being a humane man, I urge you to do it.'

'You can't foist your project on me', I protested. 'My project in life has nothing to do with all this.'

'Oh, but it has! Your project in life is, I presume, to promote the greatest happiness of the greatest number. The occasion has been thrust upon you, I admit, but the project is your own.'

'No, that is not my project. It isn't even a possible project. It is too general and abstract to be one. It is as indefinite and unhelpful as a command to do the right or good or moral thing. A project has to be particular and personal to the agent. It has to be something for a man to stake his identity on, like helping the poor on his doorstep.'

Pedro kindly refrained from pointing out that shooting an Indian would indeed help the local poor and took up the broad issue. 'Do you mean that moral questions can be settled by reference to any old project – selling vacuum cleaners, for instance – provided only that you care a lot about it? Equally, do you mean that they cannot be settled by reference to a moral principle of promoting the greatest happiness, since that cannot be part of a project? I find that very odd!'

So did I, on reflection. 'It's a matter of what can define an agent's moral identity', I repeated stubbornly.

'Naturally. But an agent can, of course, change his deepest concerns. Thus anyone whose deepest concern was with vacuum cleaners would do well to trade it in for something more justifiable. I think I know what is really at the heart of your odd line. You believe, I surmise, that every action must be prompted by a desire and that a desire to do the right thing is not a possible desire. Yet I cannot see why not. Projects, as I understand it, can be fuelled by all sorts of "desires" – for instance, dispositions of evaluation, patterns of emotional reaction, personal loyalties and all manner of commitments. Thus a judge, who has committed himself to judging solely by the law, whatever it might be and whether or not he approves of it, certainly has a project. It is a general project and one fuelled by principle rather than desire, as the term is commonly used, but one wholly allowable by the best philosophy on the subject, I fancy.'

I began to wish he had stuck to botany. 'What matters', I said, 'is that action and deliberation must always proceed from an existing motivation.' But he was not to be put off. 'Exactly,' he exclaimed, 'but "motivation" can be taken either narrowly, thus excluding dispositions, commitments and so forth, or broadly, thus including a "desire" to do the right thing. Even on the banks of the Cam, where members of the Humean race are quite common, the narrow interpretation is rare. So you are quite entitled to a project constructed on external reasons.'

I must have looked a bit blank, because he now turned fatherly. 'Look,' he said, 'you are safe enough in shooting an Indian. My government would do nothing to offend Señor Williams and even he will want you to shoot one, provided that you do not do it for utilitarian reasons. So just find yourself a non-utilitarian project, which requires you to shoot your Indian. If anyone complains, say that, since projects can hardly be self-justifying nor, once chosen, immutable, you have done your honest best to respect the requirements of morality. You won't even have to say what those requirements are, since there are several kinds of "external-reasons" theories of ethics. If you are pressed, cause a diversion by adding that utilitarianism is traditionally an "internal-reasons" theory in any case; so a utilitarian reason really ought to pass muster even on the banks of the Cam. Now, be a good chap and get on with it.'

It's funny how people can turn out to be quite nice really, when you get to know them. Anyway, reluctantly and with great distaste, I shot my Indian and was driven down to the coast in Pedro's armoured car. There are still no vacuum cleaners in that part of South America but the scandal of the execution was a factor in toppling the régime. Thus, by a bit of moral luck, I presently found myself receiving the Immanuel Kant medal of honour at the hands of the new president, Pedro himself.

*University of East Anglia,*
*Norwich NR4 7TJ*

© MARTIN HOLLIS 1983

# A NOTE ON ANCIENT ATTITUDES TOWARDS SLAVERY

*By* C. C. W. TAYLOR

IN ANALYSIS 42.2, March 1982, p. 71, Michael Slote asserts that

> ... in the ancient world – in Greece, in Rome, among the Church Fathers – no one ever wrote, or is known to have spoken, against slavery as such.

In a footnote on the same page he cites M. I. Finley, *Ancient Slavery and Modern Ideology*, pp. 120ff. as mentioning only Euripides as a possible exception to that generalization. In fact there are two clear counter-examples to Slote's generalization. The orator Alcidamas, a pupil of Gorgias, is quoted by a scholiast on Aristotle's *Rhetoric* as having said 'God has set all men free; nature has made no man a slave'. And Aristotle himself, reviewing current attitudes to slavery, writes as follows (*Politics* 1253b20-3):

> Others affirm that the rule of a master over slaves is contrary to nature, and that the distinction between slave and freeman exists by law only, and not by nature; and being an interference with nature is therefore unjust (trans. B. Jowett, in R. McKeon ed. *The Basic Works of Aristotle*, New York, 1941).

For further details see W. K. C. Guthrie, *A History of Greek Philosophy*, vol. III, Cambridge, 1969, pp. 155-60.

*Corpus Christi College*
*Oxford*

© C. C. W. TAYLOR 1983

---

# ACTING COMMITS ONE TO ETHICAL BELIEFS

*By* HARRY J. GENSLER

PRESCRIPTIVISTS commonly claim that ethical beliefs commit one to action. More precisely, the claim is that believing 'I ought to do A' logically-commits one to acting to do A. It is interesting to note that the premises used to argue for this claim would also justify a second claim, namely that action commits one to ethical beliefs. Let me sketch the argument for the first claim and then present the corresponding argument for the second claim.

## I

Let us assume, with prescriptivists, that an ought-judgment (of the appropriate evaluative sort) logically entails the corresponding imperative. Then we may argue as follows (where 'A' represents a present-tense action ('doing such and such now') and 'commits' means 'logically commits'):

1 'I ought to do A' logically entails 'Do A'.
2 If one thing logically entails a second thing, then accepting the first commits one to accepting the second.
∴ 3 Accepting 'I ought to do A' commits one to accepting 'Do A'.
4 Accepting 'I ought to do A' = believing 'I ought to do A'.
5 Accepting 'Do A' = acting to do A.
∴ 6 Believing 'I ought to do A' commits one to acting to do A.

Premise 1 follows from the prescriptivist claim that an ought-judgment entails the corresponding imperative (here 'Do A' is presumed to be addressed to oneself – some may prefer to substitute 'Let me do A' here and elsewhere). Premise 2 is analytic. Step 3 follows from 1 and 2. Premise 4 seems trivially true and premise 5 is from the prescriptivist doctrine concerning what it is to accept an imperative. The conclusion 6 follows from 3 by substituting things which are equivalent according to 4 and 5. I think the premises here are all true (although some might possibly have difficulties with 1 and 5); hence I also accept the conclusion to be true.

The conclusion here is significant. If it is true, then ethical beliefs are connected in a special way with actions – in a way different from how other kinds of beliefs are connected with actions. Let me take an example to illustrate the point. 'This man is bleeding' does not by itself entail an imperative for me to act in a given way; and if I believe that this man is bleeding, then that does not commit me, under pain of inconsistency, to doing any specific thing – the way I live cannot be inconsistent with this belief (taken by itself). But 'I ought to call the doctor' by itself does entail an imperative for me to do something ('Call the doctor'); and if I believe that I ought to call the doctor, then that does commit me, under pain of inconsistency, to doing the corresponding act – and if I don't act to do it then the way I live is inconsistent with my belief.

## II

From analogous premises, we may argue for our second claim: that acting commits one to ethical beliefs.

Consider (a) and (b):

(a) Do A.
(b) It would not be all right for me to do A.

Now (a)-(b) entails (a)-(b)-(c)-(d)-(e):

(a) Do A.
(b) It would not be all right for me to do A.
(c) It would be wrong for me to do A.
(d) I ought not to do A.
(e) Don't do A.

Here (b) entails (c), (c) entails (d), and, assuming the prescriptivist view that an ought-judgment entails the corresponding imperative, (d) entails (e). Hence (a)-(b) entails (a)-(b)-(c)-(d)-(e). But the latter is logically inconsistent; hence (a)-(b) is also logically inconsistent. So then (a) is logically inconsistent with (b). But to say that (a) is logically inconsistent with (b) is the same as saying that (a) ('Do A') logically entails (b)'s contradictory ('It would be all right for me to do A'). Hence:

'Do A' logically entails
'It would be all right for me to do A'.

Hence, if an ought-judgment entails the corresponding imperative, so also an imperative entails the corresponding permissibility-judgment. Formally speaking, this is much like saying that from '$\Box A \supset A$' one may derive '$A \supset \Diamond A$'.

This provides us with our first premise for another argument:

1′ 'Do A' logically entails 'It would be all right for me to do A'.
2 If one thing logically entails a second thing, then accepting the first commits one to accepting the second.
∴ 3 Accepting 'Do A' commits one to accepting 'It would be all right for me to do A'.
4′ Accepting 'It would be all right for me to do A' = believing 'It would be all right for me to do A'.
5 Accepting 'Do A' = acting to do A.
∴ 6′ Acting to do A commits one to believing 'It would be all right for me to do A'.

The rest of the argument is analogous to the argument given previously. Again I think the premises are all true; hence I also accept the conclusion to be true.

This conclusion also is philosophically significant. If what it claims is correct, then, contrary to what prescriptivists sometimes say, ethical judgments are not logically-optional for an acting person. Insofar as I am logically consistent (which doesn't always happen), I will both act on my ethical beliefs and also form ethical beliefs concerning my actions.

Someone might object to our second claim in the following way. 'Suppose that Justus (a morally upright person) has to act quickly without being able to gain the information needed to identify a permissible alternative. In such a case, Justus might quite consistently withhold judgment as to whether what he is doing is permissible. So he is not logically-committed to the belief that what he is doing is permissible. So your second claim seems false'.

Here we must distinguish between 'objective' and 'subjective' senses of the ethical terms. It may very well be that Justus's action is 'objectively' wrong, i.e. wrong relative to the actual facts of the situation (about which he is ignorant). But, assuming that Justus is morally upright and acting in good faith, his action is 'subjectively' permissible, i.e. permissible relative to his limited knowledge (and opportunity to reflect on this knowledge); it would be all right (subjectively) for a person who knew what Justus knew to act as Justus acted. Presumably Justus would have thought that his action was subjectively permissible but would have withheld judgment concerning whether it was objectively permissible. The example shows that we should formulate our second claim more carefully: acting logically-commits one to the belief that one's action is subjectively permissible (but not necessarily to the belief that the action is objectively permissible).

Technically speaking, we prescriptivists should claim that an ought-judgment entails the corresponding imperative and an imperative entails the corresponding permissibility-judgment, *provided that* each of these is understood in the same way (e.g. both are taken subjectively or both are taken objectively). And we should specify that 'Do A' is to be taken subjectively when we speak of accepting 'Do A' as being the same as acting to do A. So the subjective sense of 'Do A' and 'It would be all right for me to do A' must be used consistently throughout the argument.

To sum up: using standard prescriptivist principles, it would follow that a logically consistent person would not act without thinking that his action is (at least subjectively) permissible.

*Loyola University of Chicago,*
*6525 North Sheridan Road,*
*Chicago, Illinois 60626, U.S.A.*

© HARRY J. GENSLER 1983

# HARE ON UNIVERSAL PRESCRIPTIVISM AND UTILITARIANISM

*By* INGMAR PERSSON

IN his latest book, *Moral Thinking* (Clarendon Press, Oxford, 1981), R. M. Hare attempts to derive utilitarianism from his account of moral notions in terms of universalizability and prescriptivity (henceforth: UP): '. . . the formal, logical properties of the moral words . . . yield a system of moral reasoning whose conclusions have a content identical to that of a certain kind of utilitarianism' (ibid., p. 4). This is not the first time he tries his hand at this – see his *Freedom and Reason* (Oxford University Press, 1963), chap. 7 – but it is, I think, his most articulate and explicit attempt, which is why I shall here concentrate on it.

To bring off this derivation, it is essential to find a link between what one *would* prefer in certain *hypothetical* situations, when one imaginatively puts oneself into the places of other sentient beings affected by the act contemplated – this feat of imagination being what UP requires of one – and what one *does* prefer in *present* circumstances, so that this concerns the (maximal) satisfaction of the preferences duplicated – which is what utilitarianism, in the form Hare endorses it, bids one to do. Therefore, it is essential to examine what Hare has to say about the following two propositions, which he advisedly distinguishes (p. 95):

(1) 'I now prefer with strength *S* that if I were in that situation *x* should happen rather than not';
(2) 'If I were in that situation, I would prefer with strength *S* that *x* should happen rather than not'.

He asserts that though these propositions are not identical, one 'cannot know that (2), and what that would be like, without (1) being true, and that this is a conceptual truth, in the sense of "know" that moral thinking demands' (p. 96). Of this thesis Hare rightly says that it is one 'on which our whole method depends' (p. 222).

I cannot figure out what sense of 'know' Hare has in mind, for it seems to me that one can know in the fullest sense what it would be like to be in a possible situation – i.e. in the sense in which this entails imagining with great vividness and accuracy what being in this situation would be like – without it following that (1) is true. To back this up, consider the following illustration. A person, Hero, is firmly addicted to heroin, and this very moment he is seized by a strong craving for the drug. Another person, Anti-hero, has been just as firmly addicted to heroin, but has rather recently been able to break free of the habit; moreover, he regards kicking the habit as one of the best things that has happened to him. It seems to me indisputable that, in view of his own recent experience of the drug, Anti-hero can be said to know in the fullest sense what it is like (or what it would be like for him) to be in Hero's circumstances; in particular, he can adequately represent to himself the overwhelming preference for getting an injection of heroin that Hero now experiences. Yet it is surely false that this makes it necessary that Anti-hero in his present circumstances feels a preference of the same overpowering intensity that if he were in Hero's situation, he should get an injection of heroin – fortunately enough, since such a preference could hardly be conquered by another preference. On the contrary, considering his happiness with his present state of non-addiction, it is quite likely that he would now firmly prefer and want *not* to be given an injection, if in these circumstances; for he might regard this as a first step to the preferable state of non-addiction.

I think this example clearly indicates that there is something wrong with Hare's view of (1) and (2), but it does not unravel exactly what is wrong. To understand this, let us first look at the simpler case of prudence – i.e. the type of case where the only desires involved are the agent's own – and see what Hare has to say about it. Hare raises (cf. pp. 101ff) the question how preferences that the agent will have at a future time should enter into his prudential calculations at the present. Consider a person $P$ who *now* prefers that $x$ should happen *then* (sometime in the future), though he is certain that he will *then* prefer that $x$ should *not* happen *then*. (These preferences are christened 'now-for-then' and 'then-for-then' preferences respectively.) This case is not ruled out by Hare's doctrine discussed above, for the former preference could be strong enough to dominate the latter when it is reproduced in the present. What should a prudential person do in this situation, obey his now-for-then or his then-for-then preference? And what should another person $Q$ who aims to satisfy $P$'s preferences, do?

Hare attempts, though 'with a sense of leaving unfinished business' (p. 105), to deal with this dilemma by introducing a 'requirement of prudence' to the effect that 'we should always have a dominant or overriding preference now that the satisfaction of our now-for-now and then-for-then preferences should be maximized' (p. 105). To $Q$ and other persons in similar circumstances, he gives (pp. 105–6) the advice that they should take into consideration only preferences of others which comply with this requirement. $P$'s now-for-then preference for $x$ is thus disposed of as being imprudent.

I am pretty sure that Hare's treatment of this matter is fundamentally misguided. The preferences of a person are based on the representations[1] he has, representations of the nature of the item preferred in itself and of its 'consequences'. So if a person's preferences change from one time to another, this will practically always be due to changes in his representations. It is of special importance to make clear whether the change in representations consists in the person gaining further information, thus increasing the accuracy and comprehensiveness of his system of representations, or in this system becoming impoverished: the person being afflicted by amnesia or by severe pain, weariness, etc., which makes it impossible for him to recall all information he could recall in more favourable circumstances. Now, if a person now is reasonably convinced that his preference regarding $x$ will be different at a later time, he will presumably have at least an inkling as to whether this is due to his relevant representations extending or narrowing their scope. And which of these alternatives is realized is crucial to a rational or prudential person. If he believes or surmises that the representations, on which his present preference for $x$ is founded, are inferior to his future representations, on which his future preference for not-*x* relies, then he will surely revise his present preference and prefer not-*x* instead. For among the representations that determine a prudential person's preferences, there are 'meta-propositions' of the form *that there are further true propositions to be known, knowledge of which would support such and such preferences.* Hare's dilemma will then be solved by the now-for-then preference conforming to the then-for-then preference. On the other hand, if the person's conjecture is that the change in preference will be caused by a deterioration of his representations, the rational course would be to take measures to prevent this deterioration from occurring (thereby preventing the preference for not-*x* from arising), or, if this is impossible, perhaps to see to it that one does not have the opportunity of satisfying this preference when it arises. In the latter case, complying with Hare's requirement of prudence seems the very opposite of prudence: surely, it is imprudent to decide to act on preferences one believes to be based on confused or incomplete representations. For the prudential person, the then-for-then preference must here give way.

---

[1] I talk about representations rather than beliefs, since beliefs, I presume, can be dormant, i.e. possessed even at times when one fails to call them to mind. And I assume that dormant beliefs cannot generate the kind of preferences of concern here, namely occurrent preferences.

Basically, what I find strange about Hare's exposition is that he seems to conceive then-for-then preferences as immutable or unavoidable; he does not seem to recognize the possibility – indeed the necessity – of appraising and criticizing them in the light of current representations, thereby perhaps preventing them from arising. It is odd that Hare should have overlooked this possibility, in view of his being aware of the possibility of modifying current preferences by 'exposure to logic and the facts' (p. 103).

If we now revert to the discussion of (1) and (2), we see that the considerations just adduced apply here as well: one must ask whether one's representations in the hypothetical situation are superior or inferior to one's present. If the former is the case, Hare's view of (1) and (2) holds, and a preference corresponding to the hypothetical one will be formed. But on the other alternative – and assuming that the ill-founded preference cannot be prevented from occurring – the strength of one's present preference will be determined by the relation between the estimated displeasure of frustrating the hypothetical preference and the estimated disadvantage of surrendering to it.

But here there is a further bar to the inference Hare wants to draw, if the hypothetical situation is that of another person. Perhaps it is true of most individuals that they throughout their lives respond to the same representations with the same preferences. However, if we bring in other people or purely hypothetical developments of an individual into other types of personality, we must also reckon with the possibility of different preferences resulting from exactly the same representations, for presumably preferences are not fully determined by how the world is represented. This makes it possible for me to realize to the full that if I were in

*P*'s (whose representations in all relevant respects are the same as mine) circumstances, had his traits of personality (or his neurophysiological make-up), I would prefer not-*x*, but still, having the personality I in fact have, condemn this preference as, say, base, and prefer, with any intensity, it to be thwarted should it arise in me, i.e. prefer *x* to happen in the hypothetical circumstances despite the frustration this will occasion. This goes directly counter to what Hare says about the relation between (1) and (2).

This second obstacle is more difficult to surmount than the first of faulty representations. The first can be avoided by stipulating that Hare's view of (1) and (2) holds only for preferences one would have in a certain hypothetical situation *if one's representations are adequate* (by present standards). However, the second type of trouble cannot be side-stepped by some analogous stipulation such as 'and if my personality does not change', for that, of course, would render impossible the thought-experiment of putting oneself into the places of other personalities.

It might seem that Hare's view of the relation between (1) and (2) violates a principle he has committed himself to in earlier writings. In *Freedom and Reason* he affirms: 'It is, indeed, in the logical possibility of wanting *anything* (neutrally described) that the "freedom" which is alluded to in my title essentially consists' (p. 110). So, it might be argued, since a description like 'that if I were in circumstances *C*, I would prefer with strength *S* that *x* should happen rather than not' is obviously neutral in Hare's sense – i.e. it does not entail any prescriptions – it follows from the principle just quoted that one is free to adopt any preference *vis-à-vis* it. But, rather surprisingly, Hare does not regard this description as neutral; his reason for this is that it contains the term 'I' and that this 'is not a wholly descriptive word but in part prescriptive' (p. 96; cf. pp. 96–9, 221–3). Unfortunately, Hare does not point to any other ways than the inference discussed in which the alleged prescriptivity of 'I' manifests itself. Therefore, throwing in this idea does nothing to buttress the inference; it seems to be just another way of saying that the inference *is* valid.

In trying to elucidate the prescriptivity of 'I', Hare says that in identifying someone as *myself* 'I express a concern for that person' (p. 221). Even if this is correct, Hare goes wrong when he equates this concern with the acquiring of preferences that the preferences of this person be satisfied. It was out of concern for himself (and his crew) that Ulysses took precautions so that his preference to follow the seductive song of the sirens was not satisfied when it arose. It would have been fatal if his contemplating this situation beforehand had generated in him a preference of irresistible strength that when he begins to hear the song, the ship should follow it. Generally, the point is that at a moment of greater clear-sight a person can realize that the inconvenience of frustrating a desire which arises at a moment of weakness is nothing compared to the harm caused by satisfying it. Surely, it must be possible for someone to know that if or when a certain situation arises, he will, e.g. panic and react irrationally in preferring something with excessive intensity, without his forming a preference of the same strength that in this situation the first preference should be satisfied. Concern for himself can make a man prefer, with any degree of strength, the frustration of a preference imagined, of any degree of strength, to its fulfilment.

It follows from what I have said above that Hare's inference sometimes holds good, namely if the hypothetical situation in every respect that determines one's preference is identical to the present (excepting those factors which make the object of preference hypothetical in one case but categorical in the other). But to explain this we need nothing as outlandish as the prescriptivity of 'I'; the platitude that a result obtains whenever its full sufficient condition does, suffices for this.

To cap this criticism of Hare, let me note that even if everything I have so far urged against Hare is wrong, and he is right that preferences are 'contagious' through worlds, it still does not follow that his derivation of utilitarianism has succeeded. Suppose I prefer doing an action *A* which affects only two sentient beings, *P* and *Q*, who both prefer that I abstain from *A*. What UP demands of me is (a) that I first imagine myself being in one patient's position then in the other's – and consequently, if Hare is right about the contagiousness of preferences, acquire the preferences of each patient in turn – and (b) that I compare the strength of my preference with each of the replicas of *P*'s and *Q*'s preferences *individually*. If my original preference outweighs each replica *on its own*, then, on UP, I am morally justified in doing *A*. There is nothing in this requiring my preference to outweigh the *sum* of *P*'s and *Q*'s preferences or my duplicating the *sum* of these preferences. But, of course, a utilitarian would say that my doing *A* is right only if my preference – and hence my satisfaction – is more intense than the *sum* of *P*'s and *Q*'s preferences.[2] This is a problem Hare seems to slur over when he asserts that an agent can have a preference 'strong enough to outweigh all the preferences of all the others' (p. 226; cf. pp. 181-2) affected by his act – outweigh these preferences one by one or collectively?

I have tried to expose a rift in Hare's moral theory, but I am not mounting an attack on Hare for its own sake. The rift, if it exists, is important because it shows that there are *two* incompatible moral standpoints – both carrying a great deal of plausibility, I think – where Hare discerns only one. One standpoint is utilitarianism, the other is UP – the view naturally read into the Golden Rule and Kant's idea that one shall act only on maxims one can simultaneously want to elevate to universal laws – which should, I have suggested, be coupled with a doctrine of the autonomy of the will (i.e. the logical freedom of forming any preference *vis-à-vis* preferences contemplated). That utilitarianism and UP conflict was revealed in the foregoing paragraph: UP allows one to act on a preference if it outweighs the conflicting preference of each person involved *by itself*, even though it (alone or in conjunction with other similar preferences) is inferior in strength to the *sum* of the contradicting preferences. In this respect, Hare was closer to the truth in *Freedom and Reason* (see esp. chap. 9) where he recognized the possibility of a moral attitude or 'ideal' running counter to utilitarianism – though he branded it as 'fanaticism' and illustrated it with extremist examples such as Nazism – than he is in *Moral Thinking* where he maintains (pp. 171, 176–82) that such a 'fanatic' cannot exist. One should not let oneself be swayed by the condemnation implicit in the term 'fanaticism'; both of the moral viewpoints sketched have a great deal of attraction, which is why I shall not here attempt to adjudicate between them.

---

[2] Some utilitarians talk about averages of preference-satisfaction rather than sums of it, but this does not matter here.

*University of Lund,*
*Kungshuset, Lundagård, S-223 50,*
*Lund, Sweden*

© INGMAR PERSSON 1983

---

# MILL ON CONFLICTING MORAL OBLIGATIONS

*By* ROBERT W. HOAG

IN *Utilitarianism* John Stuart Mill notes that 'unequivocal cases of conflicting obligation' are 'real difficulties, the knotty points both in the theory of ethics and in the conscientious guidance of personal conduct' (II, 25).[1] Although Mill clearly holds that these cases are to be resolved by some direct appeal to utility, understanding the precise procedure for that appeal has been a 'knotty point' in the interpretation of Mill's moral philosophy.[2] I here offer an attempt to cut the knot. In the first section I rely on basic features of Mill's discussion in order to establish an approach to the interpretive issues. Then, by relying on a largely ignored example of Mill's own resolution of a moral conflict and certain general features of his moral theory, I argue that there is only one plausible interpretation of the procedure by which Mill's utilitarianism resolves cases of conflicting moral obligations.

---

[1] References to *Utilitarianism* will be provided parenthetically in the text by citing chapter and paragraph number; all references to *A System of Logic* are to the final chapter (Book VI, Chapter 12) and will be provided parenthetically in the text by citing section number following the expression 'SL'.

[2] For example: J. O. Urmson, 'The Moral Philosophy of J. S. Mill', in *Theories of Ethics*, ed. by Philippa Foot (Oxford: Oxford University Press, 1967), pp. 131–2; J. D. Mabbott, 'Interpretations of Mill's "Utilitarianism"', ibid., pp. 138–9; John Rawls, 'Two Concepts of Rules', ibid., pp. 160–1, fn.; Brian Cupples, 'A Defence of the Received Interpretation of J. S. Mill', *Australasian Journal of Philosophy* 50 (1972), pp. 133–4.

## I

Mill explicitly considers cases of conflicting moral obligations in only two places in *Utilitarianism* (II, 25 and V, 27–31). He sometimes characterizes the relevant kind of case as involving 'conflicting obligations' or 'conflicting rights and duties', and sometimes as involving conflicting moral rules, laws, maxims, or secondary principles. Furthermore, Mill appears to use these characterizations interchangeably. In my consideration of Mill's treatment of cases of moral conflict I will adopt the convention of characterizing them in terms of conflicting moral rules. Thus, for example, we can say that a moral conflict arises only if more than one moral rule is applicable to a particular circumstance and only if 'their demands are incompatible' (II, 25).

Mill clearly holds that moral conflicts are resolved by some direct appeal to considerations of utility. The need to deal with cases of conflicting moral rules is said to arise for all moral theories (II, 25), but, according to Mill, utilitarianism resolves such cases better than any other moral theory (II, 25). Mill indicates that a distinct advantage of utilitarianism consists in its permitting a direct appeal to a first principle in order to resolve moral conflicts (II, 25 and V, 30–1). Mill's first principle, the Principle of Utility, stipulates that utility is the ultimately desirable end (D. G. Brown, 'What Is Mill's Principle of Utility?', *Canadian Journal of Philosophy* 3 (1973), pp. 1–12). On Mill's view, then, a moral conflict is resolved iff the moral theory, by some procedure involving direct appeal to the Principle of Utility, specifies the morally correct behaviour for the agent in the circumstances where applicable moral rules conflict.

Given the basic structure of Mill's moral theory, we can distinguish three types of cases of conflicting moral obligations. Mill distinguishes two kinds of moral obligation: duties of perfect obligation and duties of imperfect obligation (V, 15). On Mill's view the former (unlike the latter) correlate with personal rights and are coextensive with the obligations of justice (V, 15). Given that cases of conflicting moral obligations can be characterized in terms of moral rules, we can distinguish three types of cases as follows: (1) conflicts between rules of justice, (2) conflicts between rules specifying imperfect duties, and (3) conflicts between a rule of justice and a rule specifying an imperfect duty. Mill does not explicitly distinguish among these different types of moral conflict; he only claims that the resolution of all cases of conflicting moral rules requires direct appeal to utility (II, 25). I assume that the procedure for making the direct appeal to utility is the same for all types of cases of conflicting moral rules.

Given a case of conflicting moral rules, we can distinguish three plausible procedures for resolving the conflict by direct appeal to utility. The following procedures are distinct in that the specific behaviour they would prescribe is not the same for all circumstances of moral conflict.

(I) Apply the Principle of Utility *only* to the conflicting moral rules and consider *only* the relative general-observance-utilities of those rules. An agent ought to act in conformity with the rule the general observance of which produces more happiness than the general observance of any other applicable rule.
(II) Apply the Principle of Utility directly to particular acts and consider the relative simple utilities of *all* alternative acts. The agent ought to perform the particular act which maximizes happiness.
(III) Apply the Principle of Utility directly to particular acts and consider the relative simple utilities of acts where each alternative considered is an act in conformity with at least one of the applicable rules. Given this restricted set of alternatives, the agent ought to perform the particular act which maximizes happiness.

It remains to be seen which one of these procedures is the one Mill has in mind for resolving moral conflicts by direct appeal to utility.

## II

In a largely ignored passage near the end of *Utilitarianism* Mill sketches a case of conflicting moral rules and indicates what his moral theory prescribes as a resolution of the conflict. The case is presented in order to illustrate a point about the class of perfect duties of justice:

> . . . justice is a name for certain moral requirements which . . . are of more paramount obligation than any others, though particular cases may occur in which *some other* social *duty* is so important as to over-rule any one of the general maxims of justice. (V, 37. My emphasis.)

Mill distinguishes only two kinds of moral duties, those of perfect and imperfect obligation (V, 15); therefore, on Mill's view, the class of rules of justice can be overridden by *some other duty* only if that other duty is one of imperfect obligation. Mill's point here is that the class of perfect duties can, in some particular circumstances, be overridden by imperfect duties, and he presents an example to illustrate this point:

> Thus, to save a life, it may not only be allowable, but a duty, to steal or take by force the necessary food or medicine, or to kidnap and compel to officiate the only qualified medical practitioner. (V, 37)

As the context suggests, Mill clearly presents this example as illustrating a moral conflict of type (3) above, i.e., as a conflict between a moral rule specifying an imperfect duty like the duty to render aid and rules of justice regarding property rights to food and medicine. For my purposes the importance of the passage lies in Mill's statement of the resolution of the conflict which his moral theory provides: the agent ought to steal, kidnap, and compel by force in order to save a human life.

Mill's stated resolution of his own example of a conflict between moral rules provides clear grounds for eliminating one of the appeals procedures as an adequate interpretation of Mill. Procedure (I) stipulates that the Principle of Utility is to be applied only to the conflicting rules, and that the agent ought to act in conformity with the moral rule with the highest utility. Now Mill clearly holds that the rules of justice

> concern the essentials of well-being more nearly, and are therefore of more absolute obligation, than any other rules for the guidance of life. (V, 32. Cf. V, 25, 33, 37.)

Thus, with respect to Mill's example of a moral conflict, the utility of any of the applicable rules of justice exceeds the utility of the rule specifying an imperfect duty like the duty to render aid. If Mill were committed to procedure (I), he would be committed to resolving his own example of moral conflict by specifying that the agent ought *not* to steal, kidnap, and so forth. Since Mill sees his moral theory as prescribing stealing in the circumstances, procedure (I) is not an adequate interpretation of how Mill thinks cases of conflicting moral rules are to be resolved.

Mill's example of a moral conflict makes it clear that the peculiarities of the particular case are relevant in resolving the conflict by direct appeal to the Principle of Utility. But Mill's stated resolution of his example does not indicate whether procedure (II) or procedure (III) is how the appeal to utility is to be conducted. Procedure (II) stipulates that the agent ought to perform the act which maximizes happiness in the particular circumstances. Given Mill's case where a human life is at stake, the maximally felicific alternative most likely would be to steal and kidnap in order to save a life. Similar reasoning suggests that procedure (III) would also prescribe stealing and kidnapping in the circumstances: among those acts which are in conformity with at least one of the applicable rules, the maximally felicific alternative will be acting to save a life by violating others' rights. Therefore, Mill's sole example of the resolution of a case of conflicting moral rules does not offer conclusive evidence in favour of either procedure (II) or procedure (III).

I believe that procedure (III) is the more plausible interpretation of Mill's views on resolving moral conflicts. In his brief discussions of 'unequivocal cases' of conflicting moral rules, Mill says that in such cases 'utility may be invoked to *decide between them*'. (II, 25.

Cf. V, 30. My emphasis.) That is, whatever Mill adopts as a procedure for resolving moral conflicts by direct appeal to utility, the resolution apparently involves deciding between the applicable moral rules. Procedure (II) requires the agent to maximize utility in the particular circumstances of moral conflict. In some circumstances, however, the maximal alternative may not be an act in conformity with any of the applicable moral rules; instead, maximizing utility may require performing an act which effects a compromise between the conflicting duties without being in conformity with any of the applicable rules. Therefore, procedure (II) makes no essential reference to deciding between moral rules in circumstances where they conflict. On the other hand, procedure (III) does make essential reference to the conflicting moral rules and can easily be seen as a procedure for deciding between the applicable rules. Therefore, the latter would appear to be the preferred interpretation of Mill's views on resolving moral conflicts.

This textual evidence in favour of attributing procedure (III) to Mill is not obviously conclusive: one might object, for example, that since Mill is a utilitarian, the more plausible view for him to hold is procedure (II) with its reliance on maximizing happiness in the circumstances. But some general features of Mill's moral theory add support to interpreting Mill as committed to procedure (III). First, Mill conceives of morality as a practical art which provides guidance for agents' deliberations and conduct on the ordinary occasions of everyday life (SL, §§1, 3, 6, 7; I, 2; II, 24); more specifically, his moral theory is intended to provide a *decision* procedure for 'the conscientious guidance of personal conduct' in circumstances of moral conflict (II, 25). Secondly, Mill recognizes the significant disutilities of agents' involvement in lengthy, complex calculations in the circumstances of everyday life (II, 24; SL, §3). Given the significantly smaller set of alternatives to be considered by a deliberating agent who employs procedure (III), it is at least likely that, as a decision procedure for moral agents, procedure (II) is significantly less useful than procedure (III). Therefore, attributing procedure (III) to Mill is supported by certain general features of his utilitarianism. And in the absence of evidence to the contrary, I believe that procedure (III) constitutes the most plausible interpretation of this particular feature of Mill's utilitarian moral theory.

David Lyons' recent innovative work on broader interpretive issues in Mill's moral theory also adds support to my interpretation of Mill's procedure for resolving moral conflicts. Lyons argues that Mill is a sophisticated rule-utilitarian who holds that all moral requirements are specified by some set of moral rules ('Mill's Theory of Morality', *Nous* 10 (1976), pp. 101–20; also, 'Human Rights and the General Welfare', *Philosophy and Public Affairs* 6 (1976), pp. 113–29). And, as Lyons admits, this has implications for Mill's treatment of moral conflicts ('Mill's Theory of Justice', in

*Values and Morals*, ed. by A. I. Goldman and J. Kim (Dordrecht: Reidel, 1978), pp. 1-20, espec. pp. 6-7, 10, 14). If Mill were committed to procedure (II), then, in circumstances where the maximally felicific alternative is not in conformity with any applicable moral rule, Mill's theory would be committed to the implausible view of prescribing that agents ignore all applicable moral requirements in favour of merely expedient or prudent behaviour. On the other hand, if Mill were committed to procedure (III), then the alternatives to be considered would always be limited by the applicable moral requirements, and Mill's rule-utilitarianism would never prescribe ignoring all applicable moral rules in circumstances of moral conflict. Therefore, if Mill is a rule-utilitarian, then procedure (III) is the preferred interpretation of Mill because it is the more plausible theoretical account of resolving moral conflicts. And Lyons' important interpretive work on the basic structure of Mill's moral theory then lends support to my view of how Mill's utilitarianism resolves 'unequivocal cases of conflicting obligation'.

*University of Virginia,*
*Charlottesville, Virginia 22901, U.S.A.*

© ROBERT W. HOAG 1983

---

# HUMAN REPRODUCTION: A SELF-DEFEATING STRATEGY

*By* LYNNE M. BROUGHTON

A LARGE proportion of the arguments of sociobiologists (e.g. E. O. Wilson, *On Human Nature*, Cambridge, Massachusetts, 1978; D. Symons, *The Evolution of Human Sexuality*, Oxford, 1979; R. Dawkins, *The Selfish Gene*, Oxford, 1976) is based upon a biological rationale for the double standard of sexual morality. This takes the form of a biological theory, intended as explaining why individuals invest considerable time and energy in sexual activity and issuing in the conclusion that males and females have conflicting reproductive interests. In what follows I shall be arguing that this is incoherent because it involves a set of reproductive strategies which for the human species as a whole would be self-defeating. I shall ignore the question of whether any moral standard is implied by factual matters, since if the biological theory can be shown as a failure on its own terms there is simply no question of deriving anything from it.

The theory, stated most explicitly by Wilson (op. cit., pp. 121–48), falls into two stages. The first seeks to explain two facts, *viz*:

(a) that there are species which reproduce sexually, rather than by some simpler and less energy consuming method

and

(b) that individuals of such species invest large amounts of time and energy in mating and in bringing up offspring.

The explanation of the first of these is that sex creates diversity, thus conferring on the species an evolutionary advantage. With two individuals each contributing a different combination of genes, a much greater range of genotypes appears among their offspring and there is a greater likelihood that some of the offspring may be able to survive changed environmental conditions.

The second fact is explained by postulation of an innate urge to pass on one's genes (or an urge on the part of the genes to perpetuate themselves if we are to accept Dawkins' thesis, *loc. cit.*). So each individual's investment in reproductive activities is rewarded by perpetuation of his or her genes.

But, so the second stage of the theory continues, there are significant biological differences between males and females. Females produce a relatively small number of gametes and, particularly for humans, the costs of bearing and caring for an infant are very large. Because of their relatively undeveloped state at birth (neoteny) human infants are more vulnerable at birth than those of any other species, requiring longer and more intensive nurturing for their survival. Human females during the later stages of pregnancy and in lactation have a significantly reduced ability to defend themselves from attack or even to obtain sufficient food. Hence the reproductive strategy which would best suit the female is to choose for a mate a man who can be induced to care for herself and her infant during the long period of nurturing required for enabling their offspring to reach reproductive age.

Males, on the other hand, produce millions of sperms and are (at least theoretically) capable of inseminating large numbers of females. So the best reproductive strategy for a male, on this theory, is not to spend his energy on one woman but rather to spread his genes as widely as possible. Perhaps if he does not provide care and defence for his offspring many will perish (as may some of the women with whom he mates) but there is a chance that more would survive in such circumstances than if he restricted himself to one woman.

> The resulting conflict of interest between the sexes is a property of not only human beings but also the majority of animal species... It pays males to be aggressive, hasty, fickle, and undiscriminating. In theory it is more profitable for females to be coy, to hold back ... to select males who are more likely to stay with them after insemination... Human beings obey this biological principle faithfully. (Wilson, op. cit. pp. 124–5)

It can be seen that the crux of the argument for the innateness of the double standard lies in this second stage of the biological theory which can be summarized as follows:

(i) because of their different structure and the varying parts they play in the reproductive process the strategies which would allow maximum gene transference for humans of either sex conflict

and

(ii) such a conflict is evident in all other sexually reproducing species.

Against these I shall be arguing that the suggestion of such a conflict between male and female is incoherent as a species strategy and, as a supporting point, that evidence about other species does not bear out the claim that they exhibit such a conflict.

Could a species, for which the optimal reproductive strategies of the sexes were opposed, survive? Or would it be the case that, if each individual acted optimally to perpetuate his or her genes the strategies of each would be counteracted by the behaviour of the other? Let us consider what would happen in the extreme case.

If all men and all women acted always and only on what, according to this theory, purports to be their best reproductive strategy then it would follow that, barring rape, no women would become impregnated since they would be waiting for men who would stay with them – yet there would be no such men. In order to adjust the example so that reproduction actually takes place we must postulate that there are some men and some women who do *not* act in accordance with their supposed optimum reproductive strategy. If some men are willing to stay with and care for a woman, and/or some women allow themselves to be impregnated by promiscuous males then reproduction will take place.

But what sense does this make of a proposed biological theory, by which the *optimal* reproductive strategies of men and women seem to minimize reproductive possibilities and *vice versa*.

Admittedly proponents of the biological theory would wish to hold that nothing approaching my extreme case does, or could, exist in practice. There are many other contributing factors to complicate the picture. For instance negotiations may take place between a man and a woman so that some balance of interests can be agreed upon before mating takes place. In this case each individual could be seen as compromising his or her optimal strategy to some extent in order to facilitate reproduction. Or it could be pointed out that since humans have, as part of their nature, a strong drive to seek sexual pleasure females on the whole will not be able entirely to refuse sexual contact with males.

However, these suggestions are simply complicating factors rather than a support of the theory. That negotiation may be of help in solving a conflict is not in itself evidence that such a conflict exists. That it is *needed* surely shows that the original strategy is unsatisfactory. As for the second instance, the human sexual drive is one of the facts which the biological theory purports to explain (see above para. 2) and so cannot be marshalled in support of the theory. If the suggested basic strategy is self-defeating one can envisage a range of factors which might intervene to minimize its effects. But a strategy which is self-defeating is hardly optimal.

Evidence from other species provides no support for the conflict theory. The actual behaviour of other mammals exhibits a wide range of different reproductive strategies *between* species, but for any particular species the individuals of either sex behave in highly structured and complementary (rather than conflicting) ways. There are monogamous species (e.g. wolves), promiscuous species (e.g. domestic cats), hierarchical species where all the females mate with one or a number of dominant males and weaker males lose out (e.g. walruses) and so on. But in all these species the strategies of males and females are complementary – as one might expect since reproduction does take place. After all, our theories about the biological basis for reproductive strategies are meant to explain the success of those species which survive and reproduce successfully (or the failure of those which do not).

I would contend that the conflict theory cannot account for this fact, due partly to a misunderstanding concerning the evidence upon which it is based. The vulnerability of infant and mother is taken to show that the female's optimal strategy is to find a cooperative male. But this is evidence primarily for the needs of the *infant*, if it is to survive to reproductive age, so it follows that the male's strategy in attempting to pass on his genes must be to ensure that sufficient of his progeny survive. And this, on the very evidence cited, requires him to invest time and energy in the care and defence of a small number of children and women (the number of either being proportionate to his resources). Casting large numbers of offspring upon an uncaring world would not be an optimal reproductive strategy for humans, male or female. Indeed those species for which such a strategy is optimal are characteristically those which produce large numbers of progeny capable at birth of fending for themselves.

Moreover it is not at all clear that the female's optimal strategy is to restrict the range of her sexual activity. Given that she needs male care and support during childbearing a more satisfactory strategy might be to mate with as many males as possible in order to increase the chances that at least one will stay with her. The evidence cited is not sufficient to show which of these strategies is optimal.

From the mere fact of biological difference between the sexes nothing follows concerning a conflict of interest, although some distinction in respective reproductive roles is indicated. It can only be a covert assumption that difference implies conflict which allows the theory to derive conflicting reproductive strategies. Since, as I have shown, such a conflict of strategies would be inherently self-defeating the theory postulating it fails as an explanation of human sexuality.[1]

*University of Melbourne*
*Parkville, Victoria 3052, Australia*
*and Clare Hall, Cambridge*

© LYNNE M. BROUGHTON 1983

---

[1] The final form of this paper owes much to discussion with Dr Jan Srzednicki, Dr Bernard Stonehouse and members of a graduate seminar in St. John's College, Cambridge.

---

# DICTATORS AND DEMOCRACIES

*By* LESLIE GREEN

ARROW argues that an acceptable rule for social choice should satisfy the condition of non-dictatorship: 'There is no individual whose preferences are automatically society's preferences independent of the preferences of all other individuals.' (Arrow, 1967, p. 226.) Downs (1957), Riker and Ordeshook (1973, chap. 3) and others argue that a rational individual should not participate in a social choice unless the benefits to him of participation outweigh his costs. In any large society, however, these precepts are inconsistent.

Following Arrow, assume that

(1) a social choice rule is democratic only if there exists no dictator.

Now let us consider the problem of social choice from the point of view of the economically rational individual who should

(2) participate in a social choice only if the benefits of participation outweigh the costs.

Economic man is a consequentialist who counts only his own costs and benefits. The benefits depend on the value of the individual's expressing his preferences, that is, on the efficacy of his vote. If a vote is not counted under a choice rule then its efficacy, measured as the probability ($p$) of it swinging the outcome, is zero. And in the general case, where the benefits are $B$ and the costs $C$, the value of his vote is given by $V = pB - C$. Hence, one should vote only if $pB > C$, which implies $p > C/B$. But $p$ depends crucially on the number of voters: the more there are the less likely it is that any single vote will swing the outcome. In any large electorate $p$ approaches zero. It is large only in very small electorates, and large enough to rival $C/B$ only in electorates of one.[1] But if there is an electorate of one then the efficacy of the preferences of those individuals who have no vote is zero. If one individual's preferences have a weight of one and all other's preferences weigh zero then there exists an individual whose preferences are automatically society's preferences independent of the preferences of all other individuals; and therefore the condition of non-dictatorship is not satisfied. Thus, the rational voter should

(3) participate in a social choice only if there is a dictator.

It follows trivially from (1) to (3) that the rational voter should

(4) participate in a social choice only if it is not democratic.

It seems we must abandon either (1) or (2). But Arrow calls the non-dictatorship requirement 'probably the least controversial of all' (1967, p. 226), for as a condition on the way social orderings respond to individual orderings it is a necessary one for the method of social choice to count as collective (1963, p. 30). On the other hand, to abandon (2) is to give up the economic conception of rationality as individual utility maximization. But that is a conception which structures Arrow's social choice problem in the first place. Maybe we could broaden (2) and require the individual to count *everyone's* benefits. But then the utility differential of the alternatives would have to be enormous to survive discounting by $p$. Even utilitarians want to be efficacious.

In an aside, Arrow admitted that 'the desire for a dictatorship or for a particular dictator may be overwhelming under certain conditions' (1963, p. 90). He could not have imagined that, for economically rational agents, 'certain conditions' must mean 'all conditions'.

*32 Valerie Drive, St. Catherine's*
*Ontario, Canada L2T 3G5*

© LESLIE GREEN 1983

---

[1] Those (like Downs and Riker and Ordeshook) who imagine a smaller cost-benefit ratio or who believe in probability-independent benefits can accept a weaker version of the argument. (3) becomes (3*): there is *best reason* to participate in a social choice when there is a dictator. (4) would be similarly weakened but not, I think, much less embarrassing.

## REFERENCES

Arrow, Kenneth J. 1963. *Social Choice and Individual Values* (New Haven: Yale University Press).

Arrow, Kenneth, J. 1967. 'Values and Collective Decision-Making' in P. Laslett and W. G. Runciman (eds) *Philosophy, Politics and Society*, third series (Oxford: Blackwell).

Downs, Anthony. 1957. *An Economic Theory of Democracy* (New York: Harper and Row).

Riker, William H. and Ordeshook, Peter C. 1973. *An Introduction to Positive Political Theory* (Englewood Cliffs, N.J.: Prentice-Hall).

# HUME'S SCEPTICISM CONCERNING REPORTS OF MIRACLES

### *By* ROY A. SORENSEN

IN 'On Miracles', Hume argues for an epistemic scepticism concerning miracles. Upon receiving testimony that a miracle took place, we should compare the probability of the testimony being incorrect with the probability of the occurrence of the miracle. Since the probability of incorrect testimony is always greater, a 'wise man' never believes that the miracle took place.

If this argument is sound, case by case scepticism is established. The case by case sceptic maintains that any testimony that a specific miracle took place is never sufficient evidence to merit the wise man's belief that the miracle in question took place. Given a list of reported miracles, the wise man should withhold assent from each. However, case by case scepticism does not rule out the possibility that the wise man then go on to assert that at least one of the reports is correct (though he does not know which). It may be the case that each report is improbable and yet that the probability of at least one of the reports being correct is high. The probability that a given ticket of a lottery is the winning ticket is quite low, yet one can be certain that one of the tickets is the winning ticket.

Hume appears to believe that he has established a stronger scepticism than this. Complete devastation of the argument from miracles requires the following kind of scepticism: reports of miracles are never, even collectively, sufficient evidence to merit a wise man's belief that at least one miracle took place. However, one cannot establish this kind of scepticism merely by showing that the correctness of any report of a miracle is improbable. One must show that the low probabilities of the individual testimonies do not add up in such a way as to make probable 'There is at least one miracle'.

Hume's burden of proof is heavier than he or his commentators have realized. Hume's argument does not rule out the possibility that one accepts case by case scepticism and yet one knows through testimony that at least one miracle took place.

*Michigan State University*
*East Lansing, Michigan 48823, U.S.A.*

© ROY A. SORENSEN 1983

# SELF-EXISTENCE AND THE COSMOLOGICAL ARGUMENT

*By* WILLIAM L. ROWE

IN 'Rowe, Self-Existence, And The Cosmological Argument' (ANALYSIS 42.2, March 1982, pp. 99-102), B. F. Keating raises some very interesting questions about a claim I make in Chapter Two of *Philosophy of Religion* (Belmont: Wadsworth, 1978). The claim is that The Principle of Sufficient Reason (PSR) entails the premises of the following argument:

1. Every being (that exists or ever did exist) is either a dependent being or a self-existent being.
2. Not every being can be a dependent being.

Therefore,

3. There exists a self-existent being.

(*PSR* states that there must be an explanation of any positive fact whatever. A *self-existent being* is a being whose existence is accounted for by its own nature. A *dependent being* is a being whose existence is accounted for by the causal activity of other things.)

The main point Keating endeavours to establish is that PSR alone does not entail premise (2). For the entailment to hold, he argues, we need not only PSR but also the assumption that a self-existent being is logically possible. His reasoning in support of this view is as follows. If a self-existent being is not logically possible then, given PSR, it will follow that dependent beings are the only kind of beings there could be. But if dependent beings are the only kind of beings there could be then, given the fact that there are beings, we could explain why there are dependent beings. We could explain why there are dependent beings by noting that there are beings and that there could not be any other kind of beings than dependent beings.

Suppose Keating's reasoning is correct. Suppose, that is, that if a self-existent being is not logically possible then we might be able to explain why dependent beings exist by noting (a) that there are beings, and (b) that dependent beings are the only kind of beings there could be. Keating thinks this *shows* that from PSR it doesn't follow that not every being can be a dependent being (premise 2). But he is wrong about this. For if the supposition that every being is dependent implies that there is a brute fact, a fact lacking any explanation, then that supposition is inconsistent with PSR; the truth of PSR will guarantee the truth of premise (2). To show that PSR doesn't entail (2), Keating needs to show that it is possible both that every being is dependent and that *every* positive fact has an explanation. He nowhere tries to show this. At best, he shows that if a self-existent being is not logically possible then it could be both that every being is dependent and that *some fact* – the fact that there are dependent beings – has an explanation. Moreover, in explaining this fact Keating appeals to the fact that there are beings and nowhere shows how that fact could have an explanation in the circumstances supposed. Hence, Keating has failed to establish what he set out to establish: that premise (2) does not follow from PSR alone.

Despite this, the importance of what Keating has shown should not be overlooked. If we claim, as I did, that on the supposition that every being is dependent there would be no explanation of *the fact* that there are dependent beings, we will need to suppose that a self-existent being is logically possible. And in supposing this we will be supposing more than PSR in our effort to establish that not every being can be dependent. If we continue to think, as I do, that PSR *alone* entails premise (2), we must recognize that the argument will be more complicated. In outline it would be as follows:

> If every being were dependent and a self-existent being is possible then *the fact* that there are dependent beings would lack an explanation. If every being were dependent and a self-existent being is not possible then *some positive fact or other* (perhaps the fact that there are beings) would lack an explanation.

Keating makes a second interesting point. He notes that on a reasonable characterization of the notion 'a self-existent being', if a self-existent being is possible then such a being actually exists. I have no objection to this. The point of the Cosmological Argument, however, is to establish that there exists a self-existent being of a certain sort; namely, a being that accounts for the fact that dependent beings exist. That there is a self-existent being of this sort can't be proved by arguing for its possibility.

One final point. Keating seems to grant that premise (2) follows from PSR in conjunction with the proposition that it is possible that there exists a self-existent being. He does not claim that it is not possible that there exists a self-existent being. Suppose the proposition that it is possible that a self-existent being exists is *true*. Since the proposition in question is a modal proposition, if it is true then it is necessarily true. But if it is necessarily true than (2) will follow from PSR *alone*. For if r follows from p and q and q is a necessary truth then r follows from p alone. Hence, if Keating hopes to establish that (2) doesn't follow from PSR he needs to show that it is not possible that there is a self-existent being.

*Purdue University*
*West Lafayette, Indiana 47907, U.S.A.*

© William L. Rowe 1983

# PAULSEN ON THE FREE WILL DEFENCE

*By* DAVID GORDON

DAVID L. PAULSEN in 'Divine Determinateness and the Free Will Defence' (ANALYSIS 41.3, June 1981, pp. 150–3) claims that 'one cannot consistently hold that (i) God has an eternally determinate nature; and (ii) the free will defence is an adequate response to "the problem of evil"' (p. 150). His argument proceeds by denying that John Hick and F. R. Tennant have responded effectively to a criticism of the free will defence, viz., that even if the defence is correct in claiming that a world in which persons can freely choose either good or evil is better than one in which persons are compelled to choose only good, a world in which persons always freely choose what is right is better still; thus, the actual world, in which people fail always freely to choose what is right, is not the best possible world and would not have been created by a morally perfect Being. ('[B]eing a morally perfect Being, God would have created the best possible world' (p. 150).)

Hick and Tennant argue, roughly, that a moral character acquired through struggle and in a process of development is better than a character created in a fully determined form. Tennant, as Paulsen presents his argument, goes further and claims that a character created in a fully determined form has no moral value whatever. Thus, according to Hick and Tennant, the criticism of the free will defence presented above fails. Persons who were determined always to choose the right would be less valuable than persons whose character was the result of a process of development: Hick and Tennant therefore reinstate the free will defence as a valid rejoinder to the argument from evil.

Paulsen argues that Hick and Tennant can be shown to be inconsistent if one takes account of the fact that 'Tennant, explicitly, and Hick, implicitly, affirm that God has, eternally and immutably, a morally determinate nature' (p. 152). If Tennant believes that moral goodness cannot be had *ab initio* and that God has a totally and eternally determinate nature, he cannot consistently affirm that God is morally good. Paulsen has, it seems to me, derived a contradiction here only by incorrectly representing the views of Tennant which he quotes. According to Paulsen's own citation, Tennant affirms 'moral goodness cannot be created as such; it cannot be implanted in any moral agent by an "almighty other"' (p. 152, quoting Tennant, *Philosophical Theology*, Vol. 2, p. 185). Tennant thus asserts that only for a created being is it the case that, in Paulsen's formulation, 'Moral goodness cannot be had *ab initio*'. The passage from Tennant clearly suggests that what is wrong with implanted virtue is that, because it comes from another, it cannot be ascribed to the creature possessing it. So long as Tennant accepts the non-controversial 'God is not a created being', Paulsen has failed to show a contradiction in his views.

Paulsen's success against Hick's argument is also doubtful, even if one ignores the fact that Hick apparently intends his thesis about acquired virtue to apply only to created beings (p. 151). He thinks it follows from the view he wrongly ascribes in unrestricted form to Hick, viz., 'virtue *acquired* or *developed* through a self-directing process is "richer" and "more valuable" than virtue had *ab initio*' (p. 153) that, since God presumably has not attained virtue through such a process, 'men may achieve a kind of virtue which is richer or more valuable than God's virtue' (p. 153). For this derivation to hold, one must take the major premise in the strong sense, 'Any instance of acquired or developed virtue is richer or more valuable than any instance of virtue had *ab initio*', rather than the weaker, 'Insofar as virtue is acquired or developed, it is in that respect richer or more valuable than virtue had *ab initio*'. On the latter construal, God's virtue, even though not acquired, might so surpass human virtue in other respects that Hick need not be committed to claim that men may achieve a kind of virtue which is richer or more valuable, than God's virtue. On the weaker construal of the major premise, an inconsistency would result only if Hick were to maintain 'God's virtue surpasses man's in every respect'. If Hick accepts this proposition, then Paulsen *has* shown an inconsistency in the set of propositions he sets forward. If Hick then appealed to the fact that, in the quotation Paulsen presents, he limits his thesis about acquired or developed virtue to created beings, Paulsen might well ask why the distinction does *not* apply to God. But to ask this question is not by itself sufficient to show that Hick has fallen into contradiction.

Finally, even if Paulsen's arguments against Tennant and Hick were correct, his claim at the beginning of his paper that one cannot consistently hold that '(i) God has an eternally determinate nature; and (ii) the free will defence is an adequate response to the "problem of evil"' (p. 150) is too strong. Paulsen would have shown only that Hick's and Tennant's way of response to the criticism of the free will defence is wrong. To show that one cannot hold (i) and (ii), he would have to show that any other reply to the criticism of the free will defence either fails or is inconsistent with the belief that God has an eternally determinate nature. But *prima facie*, Paulsen has not shown this, as there are many replies to the criticism of the free will defence which he does not consider. To mention one his own article suggests, 'The criticism (of the free will defence) presupposes the compatability of determinism (soft) and freedom' (p. 158). Why can't someone who denies that freedom and soft determinism are compatible consistently hold (i) and (ii)? Of course, if Paulsen could show that freedom *is* compatible with soft determinism, this line of argument would be closed. But this he has not done.

*6238 Orange Street,*
*Los Angeles, California 90048, U.S.A.*

© David Gordon 1983

# INQUIRY

*Editor: Alastair Hannay*

*Vol. 25 (1982), No. 2*

| | |
|---|---|
| The Importance of Subjectivity: An Inaugural Lecture | T.L.S. Sprigge |
| What are We Frightened of? | Barrie Falk |
| Understanding Anthropologists | Michael Luntley |
| Omniscience and Omnipotence: How They May Help — or Hurt — in a Game | Steven J. Brams |
| Frege and Kant on Geometry | Michael Dummett |
| Acting, Trying, and Essentialism (Brian O'Shaughnessy, *The Will: A Dual Aspect Theory)* | Frank Jackson |
| Text and Meaning (P.D. Juhl, *Interpretation: An Essay in the Philosophy of Literary Criticism)* | Stein Haugom Olsen |

*Vol. 25 (1982), No. 3*

| | |
|---|---|
| Grounding Rights and a Method of Reflective Equilibrium | Kai Nielsen |
| The Claims of Reflective Equilibrium | Joseph Raz |
| Self-defeating Predictions and the Fixed-point Theorem | Audun Øfsti and Dag Østerberg |
| The Role of Mathematics in the Exploration of Reality: Comments on a Paper by Herbert A. Simon | Karl Egil Aubert |
| Election Predictions: Reply | Herbert A. Simon |
| A Strong Programme for the Sociology of Belief (David Bloor, *Knowledge and Social Imagery)* | Antony Flew |
| A Paradigm for the Social Sciences? (Philippe Van Parijs, *Evolutionary Explanation in the Social Sciences: An Emerging Paradigm)* | Jon Elster |

Subscription price:
Institutions: N.kr. 150,— $27.00 — Individuals N.kr. 90,— $16.00.

Published quarterly by Universitetsforlaget, P.O. Box 2959 Tøyen, Oslo 6, Norway — P.O. Box 258, Irvington-on-Hudson, New York 10533, USA.

---

# MIND

A QUARTERLY REVIEW
OF PHILOSOPHY
edited by D. W. Hamlyn
with the assistance of S. D. Guttenplan

**Vol XCII No 365 January 1983**

**Articles**

The Necessity of Origin: H. NOONAN
Is Life Worth Reliving?: D. HEYD
Ethics, Mathematics and Relativism: J. LEAR
Agent-causation: J. BISHOP
The Difference between Dreaming and Being Awake: J.F.M. HUNTER

**Discussions**

What is Special About Democracy?: G. GRAHAM
Is Determinism Self-Refuting?: K. POPPER
Immateriality: I. THALBERG
Wiggins on Sets and Essence: G. FORBES
Socialization and Autonomy: M. BERNSTEIN

**Book Reviews**

**Books Received**

*Orders to:*
Journals Department,
Basil Blackwell Publisher Ltd.,
108 Cowley Road,
Oxford OX4 1JF, England

*Annual subscription:*
*Institutions:* £14.50 or $32.50 (Canada) $37.75
*Individuals:* £10.00 or $22.50 (Canada $26.25)
*Single issues:* £4.00 or $9.00 (Canada $10.50)

Published for the MIND ASSOCIATION by BASIL BLACKWELL PUBLISHER

# NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; T. R. Baldwin, Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. H. McDowell, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

PUBLISHER. ANALYSIS is published by Basil Blackwell Publisher Ltd, 108 Cowley Road, Oxford OX4 1JF, England. US Mailing Agent, Expediters of the Printed Word Ltd, 527 Madison Avenue, Suite 1217, New York, NY 10022, USA. Second Class postage paid at New York.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £11.00 (inland), £13.50 (overseas), $33.00 (USA), $39.50 (Canada); for individuals £7.00 (inland), £8.60 (overseas), $21.00 (USA), $25.00 (Canada). Each volume comprises four numbers, each of at least 48 pages, appearing within the academic year—in January, March, June and October. Orders should be sent to the publisher.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Mr Christopher Kirwan, Exeter College, Oxford OX1 3DP.

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

---

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by Whitstable Litho Ltd, Whitstable.

1 JUL 1983

ol. 43 No. 2 (New Series No. 198) March 19

# ANALYSIS

Edited by
CHRISTOPHER KIRWAN

CONTENTS

| | |
|---|---|
| Russell on denoting | DILIP K. BASU |
| Frege's 'beurtheilbarer Inhalt' | MARK HELME |
| On identifying reference with truth-value | GAUTAM SENGUPTA |
| Smart's argument for realism | PETER SMITH |
| Grayling on internal structure | B. J. GARRETT |
| McGinn, token physicalism, and a rejoinder of Woodfield | PAUL TAYLOR |
| Against quantum logic | JONATHAN HARRISON |
| Effective and non-effective reference | D. E. OVER |
| Vestigial modalities | GEORGE I. MAVRODES |
| Lucas, Lewis, and mechanism – one more time | ANDREW D. IRVINE |
| Personal identity and bodily continuity: a further note on 'The self and the future' | HAROLD W. NOONAN |
| Steiner's vanishing powers | JEREMY WALDRON |
| Impeccability | THOMAS V. MORRIS |

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003–2638

# Basil Blackwell

## Science and Moral Priority

Merging of Mind, Brain and Human Values

**ROGER SPERRY**

Is the human brain more than the sum of its parts? The more we learn of the way our brains work the more we begin to question whether we have any free will at all. Every reaction, emotion and thought seems traceable to some physical or electrical change in the brain. And, if we have no free will how can we speak of having values? This book, by the Nobel Prize Winner Roger Sperry, addresses these important and controversial questions.
160 pages, £12.00 (0 631 13199 X)

## Berkeley

New Edition

**G. J. WARNOCK**

'His exposition and criticism are exact, coherent and lucid.' Times Literary Supplement

'It could hardly be bettered. The book is written with clarity, conciseness, humour and precision.' The Spectator

Berkeley is one of the most influential and yet most misunderstood of eighteenth-century philosophers. In this new, revised edition of his classic introduction, G. J. Warnock examines all Berkeley's major philosophical works and discusses his most original and interesting contributions to questions still debated by philosophers today.
240 pages, hardback £15.00 (0 631 13097 7)
paperback £5.95 (0 631 13119 1)

## Morality and Language

**G. J. WARNOCK**

These essays on a number of important topics related to epistemology, morality and language are models of philosophical reasoning and lucid exposition. Among the themes examined are truth, causation, liberty and ordinary language.
240 pages, £17.50 (0 631 13098 5)

## Bradley's Logic

**ANTHONY MANSER**

Once regarded as the greatest living English philosopher, Bradley's work is now largely neglected. In this valuable reassessment of Principles of Logic Anthony Manser helps us to see how Bradley anticipated many of the problems explored by Frege and Russell, and demonstrates that it contains many of the questions which are still at the centre of philosophical discussion.
232 pages, £17.50 (0 631 13139 6)

**Basil Blackwell Publisher,** 108 Cowley Road, Oxford OX1 4JF

# Basil Blackwell

## Plato: Hippias Major

Translated with Commentary and Essay by
**PAUL WOODRUFF**

Now accepted as genuine and as a crucial text in Plato's development of the theory of forms. Hippias Major here appears in a new translation by Paul Woodruff, accompanied by an extended essay discussing the dialogue as a philosophical whole, and examining its main themes.
224 pages, £12.50 (0 631 13091 8)

## Plato's Parmenides

Translated with Analysis and edited by
**R. E. ALLEN**

Parmenides is one of Plato's most obscure, even baffling dialogues, and there is considerable controversy surrounding its meaning. This new translation and analysis offers a remarkably clear and original discussion of the work, setting it in its philosophical context and taking full account of recent Plato scholarship.
384 pages, c.£19.50 (0 631 13121 3)

## Kant's Political Philosophy

**HOWARD L. WILLIAMS**

Although recognised as an outstanding philosopher, Kant's political writings have been largely unappreciated. In this important new book, Dr Williams demonstrates that Kant's political views are not only interesting in themselves but that they are essential to an understanding of his critical philosophy as well.
(March) 320 pages, c.£20.00 (0 631 13123 X)

## Marxism and Christianity

**DENYS TURNER**

Ever since Marx described religion as the 'opiate of the people', there has been open warfare between Marxists and Christians. Communism has been accused of being amoral, while Christianity has been condemned as ahistorical. Yet, to outsiders, both seem wedded to idealism with equal fervour. Denys Turner here examines the similarities and differences between the two competing 'world-views'.
272 pages, £17.50 (0 631 13118 3)

**Basil Blackwell Publisher,** 108 Cowley Road, Oxford OX1 4JF

# RUSSELL ON DENOTING

## *By* DILIP K. BASU

THE term 'denoting complex' occurs in Russell's 'On Denoting' (all references to *OD* are to the reprint of the article in *Logic and Knowledge*, ed. R. C. Marsh, London, 1956) and has many people guessing as to what might have been Russell's meaning. Since we are not satisfied with the many guesses made, we might just make another.

We note the following interpretations of the denoting complex:

(i) A denoting complex is a denoting concept in the sense in which Russell has it in *The Principles of Mathematics* (hereafter *PM*) [1].
(ii) A denoting complex is that which denotes a meaning. It is neither a denoting phrase, nor a meaning, nor a denotation [2].
(iii) A denoting complex is identical with a sense, but Russell also speaks as though a denoting complex had a sense, a slipshod use [3].
(iv) A denoting complex corresponds to the Fregean sense [4].

All these views appear to be mistaken. Let us begin by noting that 'denoting complex' occurs in *OD* alone and only in those parts where Russell has Frege under fire. As we hope to show, Russell's use of 'denoting complex' was dictated by compulsions of the job at hand. After he had made his point Russell had no use for the denoting complex, for the denoting complex, it will be seen, forms no part of Russell's own apparatus.

## I DENOTING CONCEPTS

The doctrine of denoting concepts has been completely bypassed in *OD*. A denoting phrase is in *OD*, as it is in *PM*, a general term preceded by one of the six words 'a', 'every', 'all', 'any', 'some', or 'the', or some synonym of them. In terms of the *PM* theory the phrase 'a man' means a concept *a man* which denotes a certain combination of all men. In *OD*, however, Russell has no room for such concepts and they completely go overboard. The reason is that Russell has Frege's quantification to take care of all denoting phrases, not only of phrases prefaced with 'the'. Thus Russell declares, 'I take the notion of *variable* as fundamental' (*OD*, p. 42). He follows Frege in having the universal quantifier as *ultimate*, and by means of it takes care of 'nothing' and 'something', which, along with 'everything', are the most primitive of the denoting phrases. Now the problem[1] with 'nothing', 'everything', and 'something' is that they do not seem to come under Russell's characterization of a denoting phrase. In *OD* Russell treats 'nothing' in a proposition as he would treat 'everything', with negation transferred to the predicate. All 'no' phrases would, by this procedure now standard, lend themselves to being treated as 'every' phrases. So the problem is to see how 'everything' and 'something', which seem to behave grammatically as substantives, can be denoting phrases. 'Thing' in 'everything' is not usually brought up as a general term (Russell has use for it as a general term in *PM* §48), but in a given domain of discourse it does duty for one. This becomes clear if we consider cognate phrases such as 'everyone' and 'somewhere' which carry hints as to what general terms are involved. 'Everything' and 'something' are denoting phrases all right and Russell did not lose sight of them, for he says that '. . . a phrase containing one of the above six words always denotes' (*PM*, pp. 55–6).

---

[1] I thank the referee for pointing out this problem to me.

Matters are thus very much simplified so that we do not need different kinds of denoting concepts meant by different kinds of denoting phrases. In fact we do not need denoting concepts at all. It was left to Russell's genius to show how to let quantification take care of denoting phrases prefaced with 'the'. It has been usually supposed, and wrongly, that for Russell it is only the definite descriptions that are incomplete symbols having no meaning in themselves. Every denoting phrase shares this fate on Russell's analysis in *OD*. Thus, for instance, a proposition whose verbal expression features the denoting phrase 'a man' and which is of the general form 'C(a man)' means '"C(x) and x is human" is not always false', in which we cannot find 'a man' occurring. 'This leaves', Russell argues, '"a man" by itself wholly destitute of meaning, but gives a meaning to every proposition in whose verbal expression "a man" occurs' (*OD*, p. 43). One of the aims of *OD* was to show that this was the case with every denoting phrase. Russell states his general position clearly when he says that '*denoting phrases never have any meaning in themselves* but that every proposition in whose verbal expression they occur has a meaning' (*OD*, p. 43, our italics). Geach applies this teaching to definite descriptions alone and takes it to mean that a phrase such as 'The king of England in 1905' occurring in 'The king of England in 1905 was bearded' has no denoting concept answering to it, since the phrase as syntactical unity wilts under the pressure of Russell's analysis ([1] pp. 71–2). Since every denoting phrase embraces this fate in terms of the *OD* analysis, all denoting concepts are lost. So it is unnecessary to hark back to the days of *PM* for an interpretation of the denoting complex. It is now possible to understand why Russell's use of 'denoting phrase' excludes non-empty proper names. Such names are constants too stubborn to melt away into the variables of quantification. They do so under the prescription of Quine who wants proper names out of the way, because he does not set great store by the acquaintance of particulars. But the basic insight is the same: the main job of denoting is taken over by the variables of quantification. Russell most probably had not appreciated the power of Frege's quantification device in 1903, for any mention of Frege's treatment of generality is astonishingly absent in Appendix A of *PM*. With the ready services of quantification Russell had no use for denoting concepts, and no wonder he forgets them altogether.

## II THE DENOTING COMPLEX

To return to the denoting complex. (In the context of Russell's examination of Frege's distinction between *Sinn* and *Bedeutung* Russell's use of '*meaning*' corresponds to Frege's '*Sinn*' and that of '*denotation*' to '*Bedeutung*', and we shall follow Russell in this usage in this paper.) The interpretation that we offer of the denoting complex is this: a denoting complex is an expression with a putative meaning and a denotation.

In looking for reasons in favour of our view of the denoting complex we begin by having a look at a footnote on p. 46, treacherous according to Jager, where Russell introduces the term 'denoting complex' thus:

> Frege distinguishes the two elements of meaning and denotation everywhere, and not only in complex denoting phrases. Thus it is the *meanings* of the constituents of a denoting complex that enter into its *meaning*, not their *denotation*. In the proposition 'Mont Blanc is over 1,000 metres high', it is, according to him, the *meaning* of 'Mont Blanc', not the actual mountain, that is a constituent of the *meaning* of the proposition.

Here we have no identification, as Jager supposes, of the denoting complex and the denoting phrase. Russell, in explaining Frege's theory, makes an implied reference to the proposition 'Mont Blanc is over 1,000 metres high' as a denoting complex, and the proposition, according to Frege, has both a meaning and a denotation. So what Russell says in the footnote is in accord with the interpretation we have offered of the denoting complex.

Why should Russell need this term 'denoting complex'? In the presentation of his argument against Frege's doctrine of meaning and denotation Russell needs, to begin with, an expression in respect of which we can speak of both meaning and denotation. The denoting complex fills the bill. And when presenting his argument Russell chooses as denoting complexes expressions in respect of which it would be most plausible to distinguish between meaning and denotation, namely, definite descriptions, descriptions such as 'the centre of mass of the solar system', 'the first line of Gray's Elegy', about whose denotation we have no dispute. Russell thus avoids using in his argument proper names, function-expressions and sentences where the doctrine gets more controversial. The term 'denoting complex' was devised to serve a special purpose in *OD* and was forgotten after its due employment, for it forms no part of Russell's own theory. We therefore need not be mystified by not being able to find the term in his earlier or later writings.

If a denoting complex is as we have said it is, some remarks of Russell's which appear puzzling or objectionable cease to be so. To talk about the meaning of a denoting complex we have to mention the complex, not use it. Russell makes the point thus:

> When we wish to talk about the *meaning* of a denoting phrase as opposed to its *denotation*, the natural mode of doing so is by inverted commas. Thus we say: The centre of mass of the solar system is a point, not a denoting complex; 'The centre of mass of the solar system' is a denoting complex, not a point (*OD*, p. 48).

We find nothing objectionable here, but Jager does. His comment on this is that Russell seems to slide from 'denoting phrase' to 'denoting complex' without warning or compunction ([2] p. 55). On our interpretation of the denoting complex such sliding is most natural, the denoting phrases in question being those having both an alleged meaning and a denotation. A denoting phrase can be a denoting complex, but not all denoting phrases are denoting complexes. 'The President of India in 1981' is a denoting complex but not 'The Maharaja of India in 1981'.

One of the difficulties of the Fregean doctrine, according to Russell, is that 'we cannot succeed in both preserving the connexion of meaning and denotation and preventing them from being one'. In order to show this one naturally has to start with an expression having a (putative) meaning and a denotation. Russell therefore begins his argument very consistently thus: 'The one phrase C was to have both meaning and denotation' (*OD*, p. 49). Later in the paragraph he identifies 'C' with the denoting complex 'the first line of Gray's Elegy', thereby exemplifying the interpretation we have offered.

## III THE RELEVANCE OF RUSSELL'S THEORY OF DENOTING TO HIS THEORY OF KNOWLEDGE

It has not been always realized that the theory of denoting as developed in *OD* and carried over into his later philosophy paves the way for Russell's distinction between knowledge by acquaintance and knowledge by description. The task of denoting, according to this view, is taken over by the variables of quantification and proper names. A proper name stands for a definite object and never fails to hit the mark. This is ensured for proper names by a special kind of knowledge, namely, acquaintance. In *OD* and *Principia Mathematica* Russell accepts ordinary names such as 'Socrates' as proper names. That he later rejects their claims to being proper names, swelling thereby the ranks of denoting phrases, is a detail of his theory of acquaintance and does not represent any change in his basic insight regarding the concept of denoting. When we are operating with a variable of quantification we mean to reach out to its values, but do not get at any particular value; we choose to be blissfully indefinite. As Frege pointed out, a variable is not a name of an object, even an indefinite object, but indicates indefinitely. To put the matter in Russellian idiom, the use of a quantified variable does not put its values into a proposition and does not therefore presuppose our acquaintance with its values. As practical men we are interested, Russell tells us, in the denotation of a phrase such as 'the author of Waverley', counting as it does towards the truth or falsity of propositions in whose verbal expression such phrases figure, and the quantified variable is the means of our getting as near as possible to the denotation where there is no acquaintance to help us out [5]. Quine's wanting to eliminate all singular terms in favour of the variables of quantification is a consequence of his epistemic position regarding acquaintance with particulars. For Russell quantified variables are our refuge only where we are 'hindered by lack of acquaintance'. If we cut out the claims regarding acquaintance with particulars, Russell's position would reduce to Quine's.

Russell formulated a principle of acquaintance which Russell says has stayed with him (*My Philosophical Development*, Unwin Books edition, 1975, p. 125). The principle is that every proposition which we can understand must be composed wholly of constituents with which we are acquainted. Since a proposition involving a definite description about, say, a chair, goes over à la Russell into quantification with a variable that takes all particulars, or, at least, all material objects as its values, Sellars argues that Russell's principle of acquaintance should require that in order to understand the original proposition about the chair we must somehow be acquainted with everything. Sellars therefore concludes:

> Since it would entail that in some global sense we are acquainted with items which far transcend in number the items with which, on any occasion, we are severally acquainted, it would raise serious problems for the whole idea of acquaintance. If, as Quine argues, bound variables are the bearers of reference or, to use Russell's term, denotation, it would be difficult for Russell to avoid this consequence. Russell never seems to have seen this problem – at least in these terms [6].

Sellars' objection would be in order if a variable of quantification were a proper name of its values, which it is not, for the denotation, according to Russell, is not a constituent of the proposition except in the case of proper names [7]. But Sellars has a point; we must be able to understand the role of the variable in the reformulated proposition in a way that satisfies the principle of acquaintance. We shall let Russell speak for himself on this point:

> The first point to observe is that, in any proposition about 'the author of Waverley', provided Scott is not explicitly mentioned, the denotation itself,

> i.e. Scott, does not occur, *but only the concept of denotation, which will be represented by a variable* ([7] p. 228, our italics).

Sellars says (loc. cit.) that Russell first formulated the principle of acquaintance in *The Problems of Philosophy*. The idea is there already in *OD* (vide p. 56).

*The University of Burdwan*
*U-7/P-1, Tarabag, Burdwan-4,*
*West Bengal, India*

© DILIP K. BASU 1983

## REFERENCES

[1] P. T. Geach, 'Russell on Meaning and Denoting', *Analysis* 19.3, January 1959, pp. 69-72.
[2] R. Jager, 'Russell's Denoting Complex', *Analysis* 20.3, January 1960, pp. 53-62.
[3] J. R. Searle, 'Russell's Objections to Frege's Theory of Sense and Reference', *Analysis* 18.6, June 1958, pp. 137-43.
[4] S. Blackburn and A. Code, 'The Power of Russell's Criticism of Frege', *Analysis* 38.2, March 1978, pp. 65-77.
[5] B. Russell, 'Knowledge by Acquaintance and Knowledge by Description', *Mysticism and Logic*, Allen and Unwin (London 1917), p. 230.
[6] W. Sellars, 'Ontology and the Philosophy of Mind in Russell', *Bertrand Russell's Philosophy*, ed. G. Nakhnikian, Duckworth (London 1974), p. 59.
[7] B. Russell, *Mysticism and Logic*, p. 224.

---

# FREGE'S 'BEURTHEILBARER INHALT'

### *By* MARK HELME

IN his paper 'Frege on "Possible content of Judgement"', ANALYSIS 40.2, March 1980, pp. 83-5, Ronald Nusenhoff expresses some puzzlement about Frege's term *'beurtheilbarer Inhalt'* which I shall try to remove. He claims that in *Begriffsschrift* a judgeable content was a thought, rather than a thought plus truth value, but this appears to be incorrect. 'Content' in *Begriffsschrift* reflects not so much a conception of sense but of states of affairs – conditions of the world which obtain (*stattfinden*) or not. This is also reflected by the remark that only that which affected inferences mattered, namely the 'conceptual content', which is common to two or more expressions for a judgeable content which would both be made true by that content's obtaining (*Frege and Gödel: Two Fundamental Texts in Mathematical Logic*, ed. J. van Heijenoort, Cambridge 1970, p. 6). A judgeable content is, then, a determinate state of affairs which may obtain or not; vague sentences represent or have no such content (ibid. p. 62). Frege is operating in *Begriffsschrift* with a crude unitary conception of meaning; a sentence *'bedeutet'* a state of affairs whose obtaining would render the sentence true, and a name 'means' the object for which it stands.

In 1910 Frege wrote to Jourdain, saying that instead of *'beurtheilbarer Inhalt'* one could say *'Gedanke'*. Nusenhoff takes this as evidence that this is what Frege had meant by *'beurtheilbarer Inhalt'*, but this is better to be seen as a subsequent sophistication of *Begriffsschrift*, rather than as an explication of what Frege had meant at the time of composition. There is an inconsistency between the notions of *'Inhalt'* and *'Gedanke'* as they occur in *Begriffsschrift* and *Grundgesetze* respectively. In *Grundgesetze* I § 32 we are told that the sense of the name of a truth value (that is, a sentence) is the thought that its truth conditions are fulfilled. Now it has sometimes been said that Frege's notion of sentence-sense is to be given in terms of truth conditions, but this is not what Frege says and is in fact inconsistent with some of his other views on sense. 'The thought that...' yields an intensional context, for clearly two sentences can have the same truth conditions (when they are analytically equivalent) without expressing the same thought. The sense of a sentence is to be given in accord with the manner of its composition, and so the sense of the sentence will be determined *directly* by the sense of the constituents. Not so the truth conditions, which relate only to the possible references of the constituents. Clearly, reference for Frege is determined by sense, but no matter: Frege does not equate sense with truth conditions, and, as Dummett has argued (*Frege, Philosophy of Language*, Duckworth 1973, p. 637), Frege's claim (in a letter to Husserl, see Frege, *Philosophical and Mathematical Correspondence*, Blackwell 1980, pp. 70–1) that two contingent sentences express the same thought when they are analytically equivalent is an aberration.

When Frege says in 1892 that he has split the *beurtheilbarer Inhalt* he does not imply that he had once regarded it as having been composed out of a thought and truth value, but, having applied the sense-reference distinction to sentences, found that the reference of a sentence could not be a state of affairs but a truth value.[1]

In effect, sentences were regarded at the time of *Begriffsschrift* as names (although this was not explicitly stated), and so we cannot regard Frege's splitting of the judgeable content into thought and truth value as something which Frege incorrectly read back into the earlier work. So, I claim that Frege subsequently re-interpreted the 'contents' of *Begriffsschrift* as being thoughts, but that he did not illicitly read the notion of sentences as being names back into that book.

---

[1] 'We are therefore driven into accepting the *truth value* of a sentence as constituting its reference. By the truth value of a sentence I understand the circumstance that it is true or false'. 'Über Sinn und Bedeutung' p. 34. There is an echo of the *Begriffsschrift* doctrine of content in Frege's strange use of the word 'circumstance' here ('Umstand').

It is perhaps worthwhile to look quickly at the ways in which Frege used '*Inhalt*' after the introduction of the sense-reference distinction. The *Inhalt* of a sentence is sometimes equated with its sense (e.g. Frege, *Posthumous Writings*, Blackwell 1979, p. 167), sometimes with sense plus colouring (ibid. pp. 197–8), but the *Inhalt* of a name is always equated with its reference. Why the difference? I conjecture the following: Frege speaks of the content of a sign mostly within an arithmetical context. In 1885 in 'Formal Theories of Arithmetic' Frege compares the view that numerals are empty signs with the view that they have content. The opposite of '*inhaltlich*' would be '*inhaltslos*', i.e., 'empty', although Frege uses the word '*leer*'. This use of '*inhaltlich*' is retained when he speaks of *inhaltliche Arithmetik* (substantive or meaningful arithmetic) after 1891. In 'On Euclidean Geometry' (ibid. p. 167) Frege provides a footnote, giving as an example of the difference between a sign and a content that of numerals and numbers. The word was used most often in discussions where what was at issue was not whether there was a difference between sense and reference, but whether in the given case we could speak of such signs having a reference at all. It would not then be true to say that Frege used '*Inhalt*', at least after 1891, to mean *both* sense and reference *at the same time*. It also appears questionable to date pieces from the *Nachlass* by reference to whether '*Inhalt*' or '*Bedeutung*' was used (as was suggested by the editors of *Posthumous Writings*; see p. 72).

*University of Bielefeld,*
*Postfach 8640, 4800 Bielefeld 1,*
*W. Germany.*

© MARK HELME 1983

# ON IDENTIFYING REFERENCE WITH TRUTH-VALUE

*By* GAUTAM SENGUPTA

A FUNDAMENTAL assumption in Frege's semantics (henceforth A1) is that the customary reference of a declarative sentence is its truth-value. The purpose of this paper is to prove that A1 is false. We shall take for granted the verity of the assumption that the truth-value of a declarative sentence is a function of the references of its parts (henceforth A2). A2 is not only in conformity with Frege's view, but also entailed by Leibniz's principle. A consequence of A1 and A2 is that the truth-value of a declarative sentence containing another as part remains unchanged when the part is replaced by another sentence having the same truth-value, provided that the part as part has only customary reference and expresses a complete thought. Since we have taken the verity of A2 for granted, if the consequence is proved to be false so is A1.

Let us consider the following sentences assuming that Srimati is Bokul's youngest sister and she detests long hair:

1. Two plus two is equal to four.
2. Srimati detests long hair.
3. It is unfortunate for Ranjan that Srimati detests long hair.
4. It is unfortunate for Ranjan that Bokul's youngest sister detests long hair.

Sentence 1 is necessarily true and under the assumed circumstances sentence 2 is also true. Now; if Frege were right in assuming that the customary reference of a declarative sentence is its truth-value, then 1 and 2 would be coreferential, and substituting the one for the other in sentence 3 would have no consequence for its truth-value, provided that the embedded sentence in sentence 3 had only customary reference and expressed a complete thought. We shall first examine the nature of the reference and sense of the embedded sentence and then the consequence of the substitution in question upon the truth-value of sentence 3.

Frege admits only two kinds of reference: customary and oblique. There are three possible conditions under which the embedded sentence in question could be said not to have customary reference only. These conditions would arise in the event of the sentence having:

(i) no reference,
(ii) only oblique reference,
(iii) customary as well as oblique reference.

Since we have assumed – and our assumption is in accordance with Frege's view – that the truth-value of a declarative sentence is a function of the references of its parts, we must admit that a sentence cannot have truth-value, cannot be either true or false, if a part of it lacks reference. But given the fact that Srimati detests long hair, the fact that she does so is either unfortunate for Ranjan or it isn't. There is no third possibility. Sentence 3 is, accordingly, either true or false, i.e. it *has* truth-value. Hence the possibility of the embedded sentence in sentence 3 having no reference is ruled out.

Next, we shall consider whether the embedded sentence in question has oblique reference. If we succeed in proving that it does not have oblique reference at all, then we shall have ruled out both the possibility of its having only oblique reference and the possibility of its having both customary and oblique reference; and we shall have proved thereby that it has only customary reference.

If the embedded sentence in sentence 3 had oblique reference at all then the substitution of a part of it by an expression having the same reference but expressing a different sense would not amount to coreferential substitution and would not necessarily preserve the truth-value of sentence 3. But it does. 'Srimati' and 'Bokul's youngest sister', though coreferential, express different senses. Yet, substituting the one for the other necessarily preserves the truth-value of sentence 3. We cannot conceive of a possible world, consistent with our assumption that Srimati is identical with Bokul's youngest sister, in which sentence 3 is true but not sentence 4. We may conclude, therefore, that the embedded sentence in sentence 3 has only customary reference.

Let us consider now whether the embedded sentence in question expresses a complete thought or only part of a thought. The fact that the sense of the embedded sentence can be expressed by an independent sentence (e.g. sentence 2) having truth-value and hence expressing a complete thought indicates that it is a complete thought.

The consequence of substituting 1 for 2 in sentence 3 remains to be seen. The substitution does not necessarily preserve the truth-value of sentence 3. We can easily conceive of possible worlds in which the fact that Srimati detests long hair is unfortunate for Ranjan, but not the fact that two plus two is equal to four. A1 is thus proved to be false.[1]

*Central Institute of Indian Languages*
*Manasagangotri, Mysore 570 006, India*

© G. SENGUPTA 1983

---

[1] I am deeply indebted to Professor Michael Dummett and Professor Pranab Kumar Sen for their comments on an earlier version of this paper.

## REFERENCE

Gottlob Frege, 1892: 'On Sense and Reference' in Geach, Peter and Black, Max (eds), 1960, *Translations from the Philosophical Writings of Gottlob Frege*, Basil Blackwell, pp. 56-78.

---

# SMART'S ARGUMENT FOR REALISM

*By* PETER SMITH

REALISM, in one sense of that over-worked term, is the doctrine that our theory of the world we inhabit could be empirically ideal, satisfying all possible observational and theoretical constraints, yet still be false. In his note 'Metaphysical Realism' (ANALYSIS 42.1, January 1982, pp. 1-3) J. J. C. Smart argues for this type of realism by trying to give a particular example of how an ideal theory could consistently be supposed to be false. I will argue, however, that this alleged example fails to serve Smart's purpose.

Suppose that $T_I$ is an empirically ideal theory which agrees with modern science in at least one respect, namely by placing all physical entities in a four-dimensional space-time manifold. Smart invites us to

> . . . consider another theory, $T_I{}^*$, that is like $T_I$ in what it says about the contents of our space-time manifold, but that asserts that our space-time manifold is a cross-section of a five-dimensional manifold which contains as another cross-section another space-time manifold, also, like ours, containing galaxies, stars, planets and so on. Let us suppose that according to $T_I{}^*$ we can never know about this other space-time sub-universe because this sub-universe and our own are precluded by the laws of nature from causally interacting with one another.

Smart comments that we would never have reason to believe $T_I{}^*$ rather than $T_I$

> . . . because $T_I{}^*$ violates one of the theoretical constraints, namely that of simplicity. Nevertheless, it seems perfectly intelligible and consistent to suppose that after all $T_I{}^*$ might be true. The universe might be less simple than we hope it is, so that our theoretical constraint of simplicity leads us astray.

Now if $T_I{}^*$ is true, one crucial fact which $T_I$ actually gets wrong is the dimensionality of the physical universe. So let's begin by examining this feature of Smart's construction and ask: if the four-dimensional theory is empirically ideal, could the physical universe nevertheless be five-dimensional?

We will assume that $T_I{}^*$ locates points in the five-dimensional manifold by quintuples $(x, y, z, u, t)$; the cross-section of this space obtained by setting $u = u_0$ is our familiar space-time, while the other sub-universe which Smart postulates is situated at $u = u_s$. What we now need to know is what $T_I{}^*$ claims to be happening in other regions of the five-dimensional universe (where $u$ takes values distinct from $u_0$ and $u_s$). Smart is silent on this issue, so we need to explore a number of possibilities.

Suppose, firstly, that the five-dimensional universe is said to be quite empty apart from the objects occupying the two cross-sections at $u_0$ and $u_s$. Then all objects are entirely confined to two quite disjoint four-dimensional space-times whose contents, by hypothesis, have no causal interactions. But in this case, what sense is there in saying, as $T_I{}^*$ allegedly does, that they are physically embedded in a common unified manifold? The theory $T_I{}^*$ postulates that the physical geometry of the universe is five-dimensional – but this postulate just doesn't seem to be consistent with the concession that all objects together with their trajectories and causal relations are confined to one or another of two four-dimensional space-times, for this concession leaves nothing at all for the physical reality of the supposed fifth dimension to consist in. Suppose next that there are many space-time worlds located at different points along the $u$-axis, although again each is causally isolated from the rest. Still the same considerations apply: the claim that these worlds are physically unified into a single manifold again appears to be quite empty. (What kind of physical dimension is it in which causes do not even have contiguous effects?)

So, let's suppose, thirdly, that the various space-times *are* genuinely embedded into a five-dimensional manifold. This involves countenancing such possibilities as the following. (a) A particle can have a trajectory with a non-constant value of $u$; thus perhaps a particle can be scattered out of our space-time, where $u = u_0$, into neighbouring regions and back again. (b) Particles which are extended in our space and time can also be extended along the fifth dimension around the point $u_0$. And (c) there can be causal interactions between particles located at different values of $u$. So long as the regions around $u_0$ and around $u_s$ are sufficiently separated to be causally unconnected, these ideas could presumably be developed in a way consistent with Smart's original fantasy of two causally unrelated sub-worlds.

If $T_I$* exploits these new possibilities, then it can give explanations not available to $T_I$. To take just one example, $T_I$* could explain as the continuous path of a single particle in and out of the surface $u = u_0$ what the more restricted theory would have to regard as the separate paths of numerically distinct particles. In this way, $T_I$* might supply us with a much more powerful and successful explanatory account of the world than any four-dimensional theory. But then $T_I$ (contrary to our original hypothesis) is *not* empirically ideal. So, again, we can draw no conclusions about realism.

In short, Smart faces a dilemma. If the original four-dimensional theory $T_I$ is indeed empirically ideal, then $T_I$* can admit no causal interactions between the cross-section $u = u_0$ and neighbouring regions – which deprives the suggestion that our space-time is genuinely embedded in a physically real five-dimensional manifold of all content. But if the five-dimensional account is given real physical content by allowing that events in the cross-section $u = u_0$ are part of a wider causal nexus, then the four-dimensional theory $T_I$ must after all be less ideal than a theory that deploys the wider explanatory context available in the richer five-dimensional theory. So $T_I$ can't be both empirically ideal yet also made false by a genuinely five-dimensional theory $T_I$*.

It might be objected, however, that I have challenged only an inessential feature of Smart's case for realism:[1] for the key idea underlying his argument is simply that there might be two sub-universes which are causally unrelated (and the supplementary suggestion that these sub-universes might still be related geometrically is a dispensable refinement). But why should it be thought that this key idea supports the kind of realism for which Smart says he is arguing? Even if we grant that the idea of there being sub-universes unconnected with our own makes sense, this certainly doesn't threaten the status of $T_I$ as a true account of our portion of the world. In other words, we might grant that (I) our empirically ideal theory $T_I$ may not tell the *whole* truth about the world because there are, or could be, empirically inaccessible truths. But the conclusion which Smart wanted was that (II) our empirically ideal theory might in fact be false. And (I) does not entail (II) unless one supposes that $T_I$ involves a clause to the effect that it does tell the whole truth – and why should one suppose *that*? Surely one feature that we want our theories in general to possess is what we might call lack of surplus content: If $T_1$ and $T_2$ are observationally equivalent, $T_1 \subset T_2$, and $T_1$ is just as manageable, elegant, etc., as $T_2$, then we prefer $T_1$. In particular, our ideal theory $T_I$ should lack surplus content. But the addition to a theory which already meets all observational and theoretical constraints of a clause to the effect that there are no further truths not captured by the theory is precisely the addition of some empirically unconstrained surplus content. In fact our ideal theory $T_I$ should be left open-ended, and hence the speculation that it doesn't tell the *whole* truth cannot be used to falsify $T_I$.

---

[1] Professor Smart himself raised this objection: I am most grateful to him for an extensive and illuminating correspondence.

Of course, we don't need to resort to fantasies about duplicate sub-universes to allow that there can be empirically inaccessible regions of the world (just suppose a spatially infinite universe of finite age, where – as in our world – causal influences propagate with only finite speed). Then couldn't it be argued that our theory $T_I$ might be empirically ideal yet falsified by happenings in the inaccessible region of the universe? But again, why must the opponent of realism require an ideal theory to address itself to what, by its own lights, has no possible causal contact with us? Why not insist that for any theory T, the theory T′ obtained by restricting T to what it deems to be the causally accessible part of the universe is nearer to being empirically ideal, just by having less surplus content which floats free from any possible observational constraint (cf. Quine's maxim 'Don't venture further from the sensory evidence than you need to'). But then the ideal theory, being subject to such a restriction, will not be vulnerable to falsification by inaccessible happenings.

In short, if Smart's sub-universes are left geometrically unrelated, then his speculations cannot support the claim that our ideal theory may yet be false within its intended domain: on the other hand, his original story which embedded the sub-universes in a physically real five-dimensional manifold would indeed falsify any four-dimensional theory of the world but is inconsistent with the supposition that our four-dimensional theory $T_I$ is ideal. Hence neither version of Smart's argumentative strategy succeeds.[2]

*University College of Wales,*
*Aberystwyth, Dyfed SY23 3DY*

© PETER SMITH 1983

---

[2] Thanks are due to the referee, who suggested a number of neat stylistic improvements to my initial draft of this note.

---

# GRAYLING ON INTERNAL STRUCTURE

*By* B. J. GARRETT

A. C. Grayling ('Internal Structure and Essence', ANALYSIS 42.3, June 1982, pp. 139–40) has recently put forward a number of arguments which, he claims, show that the conception of internal structure as the essence of natural kinds (the line taken by Kripke and Putnam) does not make sense. In this note I attempt to show how his arguments fail.

Grayling first points out that internal structures are nested, that there are different levels of internal structure (e.g. physiological structure, cell structure, molecular structure, atomic structure and so on). Which level of internal structure, asks Grayling, is supposed to be *the* level constituting the essence of a given natural kind?

To begin with, it is unclear that talk of essence comes to anything more than talk of essential properties – to say that it is part of the essence of some individual *a* that it is F is surely just to say that there is no possible world in which *a* exists and lacks the property of being F, i.e. that being F is an essential property of *a*. Under such a construal the essence of *a* is the conjunction of *a*'s essential properties.

Accordingly, the essence of any given natural kind consists in the conjunction of properties essential to that kind. But why should it be assumed that all the structural properties essential to a kind cluster around just one level of internal structure? Does an essentialist position on the internal structure of natural kinds entail that there be only one level which constitutes the essence? If it does, then Grayling's scepticism concerning the notion of internal structure as essence would appear to be in order.

Surely, however, it does not. For an essentialist could argue that *all* levels of internal structure are essential to members of a kind – if kinds *K* and *K'* are such that members of *K* possess different internal structures from members of *K'* at any level(s) then *K* and *K'* are distinct kinds. There is indeed no one level which alone constitutes the essence of a kind. But such a conclusion does not thereby impugn the legitimacy of the concepts of essence or essential property when applied to natural kinds.

Grayling then argues that if differences of internal structure mark differences of kind, then we get the absurd result that every individual is its own kind since every individual is unique and such uniqueness stems from differences in internal structure. However, this argument is unsound (and even if it were sound it would presumably only apply to biological natural kinds and not to non-biological natural kinds such as gold and water). For it is not just *any* difference in internal structure which is relevant to difference in kind. It is not differences within the same structure at a given level (i.e. differences which account for divergences in external structure between members of the same kind) which are relevant to kind-differentiation but differences of structure (i.e. different structures) at a given level. Consequently, the fact that, to use Grayling's example, a particular tiger has longer or shorter hairs than another tiger does not entail that each individual tiger is its own kind, since such differences do not stem from differences of structure but only from differences within the same structure (possessed by all tigers) at the relevant levels.

Grayling's final argument meets with a similar fate to the above. We possess criteria for sorting things into kinds. But if it is logically possible that internal structures be infinitely nested and we have no way of choosing the level of internal structure essential to a kind then, argues Grayling, such criteria can have nothing to do with internal structure. However, from what has been said above, it is clear where the fault in this argument lies. For given that there is no one specific level of internal structure which constitutes the essence of a kind, and that we have access to some (essential) levels, then it clearly does not follow that kind criteria have nothing to do with internal structure – if all levels are equally essential then access to some kind-discriminating level of internal structure will provide criteria for kind-membership.

It is important to note that the 'internal structure' essentialist is only committed to the view that the satisfaction of such criteria (and indeed correspondence at all levels) is a necessary condition for membership of a kind. Whether structural correspondence at all levels provides a sufficient condition for kind-membership will depend upon whether the kind in question is biological or non-biological. In the latter case, considerations of internal structure may well be sufficient but in the former case such considerations will be sufficient only if essentialist claims concerning common genetic origin are false.

Grayling's arguments, therefore, are not telling against an essentialist position. His basic mistake, I have suggested, was to assume that if there is not one and only one level of internal structure which constitutes the essence of a kind, then the notion of internal structure as essence is unintelligible. But Grayling has given no reason why the essentialist cannot rest content with the idea that all levels of internal structure are essential to (i.e. partially or wholly constitute the essence of) a natural kind. Further, if all Grayling's arguments went through (even in conjunction with arguments against the necessity of causal origin and the necessity of common genetic origin) then it certainly would not follow that we ought to 'eschew talk of essence altogether' – for, after all, the most persuasive forms of essentialism do not rely upon considerations of either origin or internal structure.

*Department of Logic and Metaphysics,*
*The University, St. Andrews,*
*Fife KY16 9AL.*

© B. J. GARRETT 1983

---

# McGINN, TOKEN PHYSICALISM, AND A REJOINDER OF WOODFIELD

*By* PAUL TAYLOR

KRIPKE'S objection to Physicalism explicitly places under attack both Type and Token versions of the Identity Thesis. Thus, included in Kripke's discussion in 'Identity and Necessity' (in M. Munitz (ed.) *Identity and Individuation*, New York, 1971) is the example: 'My being in pain at such and such a time is my being in such and such a brain state at such and such a time.' But Colin McGinn has argued (ANALYSIS, 37.2, March 1977, pp. 78–80) that the Token Identity Thesis, unlike the Type Identity view, is not vulnerable to the objection. McGinn reasons as follows. The Cartesian intuition that pains could occur without the accompanying firing of C-fibres damages the claim that, necessarily, pains (in general) are C-fibre firings. But how does this intuition affect the Token Identity Thesis? The latter says that each particular pain state is identical with some or other physical state, and if Kripke is correct, these identities are necessary if true at all. Thus if my pain now is the same as the firing of certain C-fibres in my brain, then the pain is necessarily that firing of C-fibres. Now McGinn points out that the mentioned Cartesian intuition, that *some or other* pain could occur without any C-fibres firing at all, is perfectly compatible with the claim that my pain now is necessarily the firing of certain C-fibres; indeed, as he says, a Token Identity Theorist is likely to admit gladly that many other pains are more than likely *not* realized in the way mine is now. Thus, according to McGinn, the Token Identity Thesis is unaffected by the Cartesian intuition, and is therefore invulnerable to Kripke's objection.

I find this defence unconvincing. For surely the relevant Cartesian intuition in the case of token identities will be something like this: 'Suppose that the pain I am now having is realized by a presently occurring firing of C-fibres; well, *that* pain could have occurred without *that* firing of C-fibres.' (cf. Peacocke, *Holistic Explanation*, p. 133). If this intuition is allowed to stand then, of course, it damages the Token Identity Thesis. I admit that this version of the Cartesian intuition does not sound as compelling as the one pertaining to types, but I think that anyone who accepts the spirit of Kripke's argument for dualism, and even those who don't but reject the Type Identity view – say, because of the alleged variable realization of the mental – should also reject the Token Identity Thesis. My reason is this. A certain kind of Cartesian (and apparently Kripke too) believes that pain has a phenomenal essence. It follows from this view that the existence and persistence of any particular pain will depend crucially on the fact that a certain phenomenal state exists. Some Functionalists, on the other hand, have argued that pain is essentially a functional state, and it follows from this view that whether a particular pain now exists and persists will turn crucially on the question of whether a certain functional property is instantiated. Both the phenomenalist and functionalist accounts of pain are compatible with (though neither, I think, entails) the idea that pains are physically realized, but I believe *both* views provide a problem for the theorist who wishes to claim a strict token identity between particular physical states and the occurrences of pain they realize. Suppose I have a pain which lasts for a few minutes and then goes away, and suppose also that the realization of this pain was transferred from one set of C-fibres to another after the first few seconds of its onset. I think the Token Identity Theorist will now say that in that case the pain is identical to a complex physical state including the *series* of firing C-fibres. But now I come to what I referred to earlier as the 'spirit' of the Cartesian view. The Cartesian, if he accepts the idea of realization at all, will hardly accept that whatever state happened to realize my pain *necessarily* realized that pain. And here he is on firm ground. For suppose I am kicked, it hurts, and there is a firing of C-fibres. The Cartesian intuition here seems very compelling: it is that the pain resulting from the kick would have been the same pain even if, rather than being realized by the one set of C-fibres firing, its realization had been 'transferred' after a few seconds. A lurking functionalist intuition is equally compelling. The functionalist belief in the 'plasticity' or variable realization of mental *kinds* will tend to prevail when mental particulars are considered. A particular mental state, *whether or not* the realization of that state is transferred inside the brain in the way I have imagined possible, will remain one and the same state, providing only that its origin and overall causal effect on the manifest behaviour of its possessor remains the same. But then, no particular realization that does occur *had* to be the realization of that particular mental state; the state just happened to be realized in the way it was, and a slightly different history of realization would not have affected its identity.

So if we accept Kripke's view that identities like this must be necessarily true if true at all, we must, it seems, reject the Token Identity Thesis.

Andrew Woodfield (ANALYSIS, 38.3, June 1978, pp. 140–3) has also objected to McGinn's defence of the Token Identity Theory, but has suggested another defence, which is also a defence of the Type Identity Thesis. He suggests that instead of looking for an epistemic counterpart of pain, we consider the other side of the identity equation. We can explain away the seeming contingency of 'Pain is the firing of C-fibres', he thinks, by adducing the fact that our means of telling the presence of C-fibre firings could be such that the evidence we regard as sufficient for inferring their presence is compatible with their absence and with the presence of some other neuronal state which affects our observation instruments in just the same way. Woodfield takes this consideration as sufficient for disengaging the tenacious Cartesian intuition. (Another version of this defence of the Identity Theory can be found in Richard Boyd, 'Materialism without Reduction: What Physicalism Does Not Entail', in Ned Block (ed.) *Readings in Philosophy of Psychology*, Vol. One (Harvard University Press, 1980). McGinn's article is reprinted in this collection.) Again I find the defence unconvincing.

To begin with, the discovery of epistemic counterparts on one or other side of a candidate identity claim which is such that, if true, it must be true necessarily, will not always free that claim of intuitive implausibility. I dare say we could find epistemic counterparts for umbrellas, or bats, but still, we know intuitively that umbrellas aren't the same as bats. All we need, to know this, is a general idea of what kinds of things we're concerned with here. The conceivability of our usual evidence for the presence of bats being equally good evidence for something else is irrelevant.

But, of course, the claims 'Heat is molecular motion' and 'Pain is the firing of C-fibres' are more plausible identity claims, even though both have a contingent ring. In considering claims like these, Kripke uses the idea of an epistemic counterpart (though the latter phrase is not his) to establish whether our usual means of telling the presence of something designated by a given term occurring in an identity claim involves the awareness of a feature which finally determines the designation of the term. He finds that we usually tell of the presence of what we call 'heat' by establishing the presence of a feature which is in fact compatible with the absence of heat, namely its disposition to affect our thermal sense in a certain way. We can conclude that this disposition is not an essential property of heat, and 'heat' does not designate rigidly in virtue of that property. Now we know *in advance* that 'molecular motion' picks out a kind in virtue of a property distinct from that disposition, but since the latter is not an essential property of what 'heat' refers to, it remains an open possibility that 'heat' and 'molecular motion' refer to the same thing in all possible worlds.

Now consider the claim 'Pain is the firing of C-fibres'. We usually tell the presence of pain by our awareness of the presence of a certain sensation. Is our awareness of that sensation a contingent accompaniment to the presence of pain? Kripke says no; the fact that pain has no epistemic counterpart – the fact that if *that* awareness is present then pain is present – tells us that what we are aware of here is the very essence of pain. Looking for an epistemic counterpart reveals that 'pain' refers to something that feels like *that* in all possible worlds. But we know *in advance*, whatever our observational problems or ignorance about the brain, that the designation of the term 'firing of C-fibres' is finally determined by some or other physical property; therefore *not* a feeling or phenomenal property. And that is enough to show that 'pain' and 'the firing of C-fibres' do not denote the same type in all possible worlds. Once we know that these terms designate their objects in virtue of different essential properties, the possibility of epistemic counterparts to the firing of C-fibres is no longer of consequence.[1]

*University of Cape Town,*
*Rondesbosch 7700,*
*South Africa.*

© PAUL TAYLOR 1983

---

[1] I would like to thank Denise Meyerson and Lawrie Reznek for their helpful comments on the issues I raise in this paper.

---

# AGAINST QUANTUM LOGIC

*By* JONATHAN HARRISON

IT is generally supposed that quantum theory gives rise to a difficulty for classical logic. If a particle cannot have a determinate position and velocity at the same time, then, though the following proposition

(1) $p0 \cdot (v0 \vee v1 \vee \ldots \vee vn \vee \ldots)$

is true, the proposition

(2) $(p0 \cdot v0) \vee (p0 \cdot v1) \vee \ldots \vee (p0 \cdot vn) \vee \ldots$

is false. Proposition (1), therefore, cannot imply proposition (2). This means, at the very least, that quantum theory implies that we must reject the principle of distribution: $(p \cdot q \vee r) \supset ((p \cdot q) \vee (p \cdot r))$. Hence 'logics' (quantum logics) must, and have been, evolved which do not contain this principle. To give just one example, Professor Putnam has devoted a seminal article[1] to discussing why and what modifications must be made to classical logic to accommodate quantum theory. More recently a whole symposium has been devoted to the subject at the 1982 Joint Session of the Mind Association and Aristotelian Society.[2] There is much other literature.

There would be no reason to suppose that (1) did not imply (2), and so no reason to reject the principle of distribution and evolve quantum logic, if, in the event of quantum theory's being true, (1) and (2) were *both* false. And I had always supposed that, according to quantum theory, (1)'s second conjunct not only could be false, but would be false if its first conjunct were true. Given that a particle's position is determinate, it is false that it has any of the velocities specified in the second conjunct of (1). Hence, if quantum theory is true, the truth of the first conjunct in (1) implies the falsity of the second, and (1) itself must be false. Thus the very circumstance, that a particle cannot have a determinate position and velocity, which implies the falsity of (2), also implies the falsity of (1), and the difficulty for classical logic is removed.

Professor Putnam, among others, appears to think that (1) must be true, given the truth of its first conjunct, because he has supposed that its second conjunct is a logical truth. In his words: 'Since the "span" of Vr1, ..., VrN is the whole space, the statement: $m(s) = r1 \vee m(s) = r2 \vee \ldots \vee m(s) = rn$ (where m is the magnitude in question) is an always-true statement.' (I do not think my easier mathematics makes any difference of principle.) But if the second conjunct in (1) is a logical truth, quantum theory must be false, for quantum theory just asserts that a particle does not have to have velocity v0, or velocity v1, etc., for all the possible velocities there are. Contrapositively, if quantum theory is true, the second conjunct in (1) cannot be a logical truth.

The second conjunct in (1) has probably *seemed* to many philosophers, including Professor Putnam, to be a logical truth because v0, v1, etc., are all the possible *velocities* there are. And that these are all the possible velocities that there are has been confused with the statement that the *propositions* asserting that a particle is at v0, that it is at v1, etc., are all the *possibilities* (possibly true propositions) that there are. The proposition that these are all the possibilities that there are must be assumed by those who wish to reject classical logic, for otherwise the second conjunct in (1), like numerous other disjunctive propositions, can be false simply because it does *not* exhaust the possibilities.

---

[1] 'The Logic of Quantum Mechanics', in *Philosophical Papers*, Vol. 1, C.U.P., 1975. This article was first published as 'Is Logic Empirical?' in *Boston Studies in the Philosophy of Science* 5, edited by R. Cohen and M. Wartofsky, D. Reidel, 1968.

[2] '"Or", "Not" and the Way Things Are'. The symposiasts were Peter Gibbins and W. Newton-Smith.

A possible velocity, however, is not the same thing as a possibility, so the second conjunct in (1) does *not* exhaust the possibilities. The additional possibility is that the particle is not travelling with any of these velocities. The appearance that the second conjunct in (1) is a necessary truth should be dispelled once it is realized that to say that these are all the possible velocities that there are simply means that a particle must be travelling at one or other of them, *if it has a determinate velocity at all.*

Either the second conjunct in (1) is logically true or it is not. If it is logically true, quantum theory must be false, so there is no need to revise classical logic to accommodate it. On the other hand, if it is not logically true, there is no reason why it should not be false in the event of quantum theory's being true, and so (1) and (2) be both false in the same circumstances. Hence quantum theory provides no reason for rejecting the principle of distribution, and there is not this need for quantum logic. Whether there is any need for quantum logic at all I have not the technical competence to say, but I doubt it.

*University of Nottingham,*
*University Park, Nottingham NG7 2RD*

© JONATHAN HARRISON 1983

# EFFECTIVE AND NON-EFFECTIVE REFERENCE

*By* D. E. OVER

DONNELLAN (1966) introduces his well known distinction between attributive and referential uses of definite descriptions in these words (p. 285):

> A speaker who uses a definite description attributively in an assertion states something about whoever or whatever is the so-and-so. A speaker who uses a definite description referentially in an assertion, on the other hand, uses the description to enable his audience to pick out whom or what he is talking about and states something about that person or thing.

Donnellan's prime example concerns a lovable man, Smith, who has been brutally murdered. One of Smith's friends might discover his mutilated corpse sometime after the murder and infer, 'Smith's murderer, whoever he is, is insane.' This use of 'Smith's murderer' would be called attributive by Donnellan. But suppose that Smith's friends actually witness the murder and catch the hysterical murderer. One of them might simply state, 'Smith's murderer is insane.' Donnellan would call this use of 'Smith's murderer' referential. Now unfortunately, intriguing examples of this type do not themselves explain what the 'whoever' (or 'whatever') qualification means, nor do they explain what it is for a speaker (and his audience) to be able to pick out whom (or what) he is talking about.

The way to give these explanations, it seems to me, is to use the distinction between effective and non-effective reference. These concepts have not received the attention they deserve, though they are essentially connected with the much discussed distinction between constructive and non-constructive aspects of natural language, and with what Dummett (1973) calls an effective language (p. 230). A speaker uses a singular term with effective reference if he associates with it an effective method of deciding whether any given object is its referent; otherwise the speaker uses the term with non-effective reference. The connection with constructivity is immediate. A speaker can justify an assertion constructively only if he uses the singular terms in the assertion with effective reference. If he uses the terms with non-effective reference, then he can only give general, non-constructive grounds for his assertion.

Donnellan's main example brings out the point quite well. The first possibility above was that the speaker did not see the murder and did not know who was responsible. He believed that the murderer was insane on the general grounds that anyone who could act in so brutal a way must be insane. Donnellan comments (p. 291):

> The presumption that the speaker believes *someone* is Smith's murderer does not arrive here from a more specific presumption that he believes Jones or Robinson or someone else who he can name or identify is Smith's murderer.

In effect, Donnellan is making the point that the speaker who uses 'Smith's murderer' attributively can only justify non-constructively his belief that some single person murdered Smith and is insane. A constructive justification for a statement of the form $\exists xA(x)$ must effectively refer to a particular object which satisfies $A(x)$. But the first speaker cannot do this; he can only use 'Smith's murderer' with non-effective reference. In the second case above, however, the speaker saw the murder and the murderer's mad behaviour; he can effectively refer to the murderer and justify his belief constructively.

Not all Donnellan's statements about the attributive/referential distinction are completely clear or even consistent, but an attempt can be made to capture the essence of his position. A speaker who uses a description referentially must have a certain kind of ability. (His ability is surely more important for the distinction than his audience's.) This ability is to be explained in terms of the speaker's possession of an effective method for deciding which given object is the referent of the description he uses. If the speaker has no such method, then he uses his description attributively, has no particular object in mind, and must talk about whoever or whatever satisfies the description. It is often said that the 'whoever' or 'whatever' qualification is itself ambiguous. It is used at times by speakers to show that they do not know the name, the occupation, the scientific classification, the function, or some other interesting fact about an object which they can effectively pick out. But the phrase can also be used to indicate that a speaker does not have an effective way of picking out the object referred to. Given a group of men at a police identity parade, a speaker could say without inconsistency, 'I know who the murderer is – he is *that* man, whoever he is'. And another speaker, after failing to point anyone out, could say, 'Well, the murderer is a brute, whoever he is.' It is this latter kind of use, which can be clear in context, of the 'whoever' or 'whatever' qualification that is a sign of the attributive use of a description.

Dummett (1973) points out, in his account of an effective language, how necessary it is to explain what it means to be *given* objects to which an effective method can be applied. It is at first plausible, he holds, that the 'favoured method' of being given an ordinary object should be '. . . by the use of a demonstrative accompanied by a pointing gesture' (p. 232). Donnellan's examples of his distinction seem to be built on this particular 'favoured method'. For examples of the referential use, he asks us to consider objects which could be 'given' to a speaker by pointing, such as men on public trial or people at a party. The speaker's ability to refer effectively would be displayed by his choice of one of the given objects; he can, in the example of Smith's murder, choose which trial to go to and whom to point out there (or at least he can point to a photograph of the man he intends to refer to in a newspaper). With the attributive examples, either the speaker cannot be 'given' the object among others at all (the murderer has escaped into the population at large), or he cannot point it out when it is included in a 'given' group (as in a police identity parade).

Donnellan originally made much of a supposed difference between attributive and referential uses of a description when the speaker made some mistake about whom or what he was referring to. Fortunately our analysis shows that his distinction can be drawn, especially in his favoured examples, without mentioning mistakes, as his detailed claims about them have been strongly criticized. (see Kripke (1977).) He holds that a speaker can 'say something true' in uttering 'Smith's murderer is insane', even if Smith has not in fact been murdered. The speaker could do this, so Donnellan argues, if he has noticed the insane behaviour of Smith's accused murderer in court, and uses 'Smith's murderer' referentially to enable his audience to pick out this accused man. Donnellan at first held that, under conditions like these, a speaker could not 'say something true' if he tried to use the description attributively to refer to whoever murdered Smith, as he falsely believed, in a brutal way. Later Donnellan (1968) came to accept that a speaker could 'score a near miss' with an attributive use of a description and still 'say something true', but he continued to argue that only with a referential use could a speaker 'miss by a mile' and still 'say something true' (p. 209).

There is obviously a problem in interpreting Donnellan's vague statements about mistakes. Depending on our theory of descriptions, we could hold either that the sentence uttered by a mistaken speaker was false or that it had no truth value. It is hard to see how we could justify the claim that it was literally true. (Kripke (1977) makes this point.) However, there does seem to be some justification for what Donnellan has said. A speaker could associate with the description in 'Smith's murderer is insane' an effective method for picking out Smith's accused murderer, Jones, who acts insane in court. The speaker's audience could grasp from the context that this was his method, and so grasp that he intended to make a statement about Jones, *viz.* that he is insane. This *intended* statement could be true and constructively justified by Jones's behaviour, even if the story of Smith's murder was completely fabricated. Imagine now that there is still a story of Smith's supposedly brutal murder, but no one is accused of the crime. A speaker would try to use his description attributively and with non-effective reference in saying, 'Smith's murderer, whoever he is, is insane.' His only way of trying to refer is by means of the description he actually uses or a closely related one from the story, such as 'The man who attacked Smith'. If the story about Smith is completely groundless, and he has not been murdered or even attacked, then there is no way of formulating a true statement about someone from the speaker's state of mind. But there might be a way of doing this if Smith has at least been attacked; the speaker could then be seen as holding, with non-constructive justification, that the attacker of Smith, whoever he was, was insane. A truth can only be extracted from this speaker's non-constructive reasoning if he 'scores a near miss'; the former speaker can 'miss by a mile' and his intended statement still be true, since his constructive reasoning is grounded in a particular person, Jones, who does exist and may be insane. But it should still be stressed that, in both cases, the actual sentence uttered is false or has no truth value.

The possibility of a mistake could, of course, make a speaker careful about how he justifies his assertion. Suppose that a speaker has *both* a constructive and a non-constructive way of justifying his statement that Smith's murderer is insane. He could use 'Smith's murderer' with effective reference, and constructively justify his statement by pointing out the insane behaviour of the murderer (or at least of the man he believes to be the murderer) in court. Alternatively, the speaker could ignore the supposed murderer and his behaviour, and give general non-constructive grounds for holding that anyone who could so horribly murder Smith must be insane. He might use the latter justification because he had some doubts that the man in court was the murderer. Donnellan (1966) states that '. . . whether or not a definite description was used referentially or attributively is a function of the speaker's intentions in a particular case.' (P. 297) Yet the speaker cannot show that he has one of these intentions rather than the other by construing or phrasing his utterance one way rather than another, as Donnellan holds that this distinction is not one of syntactic or semantic ambiguity. Donnellan does not state how a speaker can display his intention, but our interpretation of his position can do this. A speaker intends to use a description referentially in an assertion if he tries to justify the assertion constructively, by at least openly associating with the description what he takes to be an effective method for picking out its referent. And a speaker intends to use a description attributively in an assertion if he tries to justify the assertion non-constructively. The difference is one of epistemology.

Kripke (1977) points out that Donnellan (1966) does not clearly distinguish between speaker's reference and semantic reference. We should not make the same mistake. *Speakers* use descriptions attributively, referentially, with effective references, or without it; and these concepts are to be explained by whether or not the speakers have an effective method for deciding whether any given object is the *semantic* referent of the description. The semantic reference of a description is fixed by its meaning, but a speaker associates with a description in a referential use an effective method that may not have a necessary connection with its meaning. The speaker, of course, thinks (or expects his audience to think) that the method does, as a matter of fact, lead to the semantic referent of the description, but he may be mistaken. Even if he is mistaken, he still has an *intended* referent, given by the effective method he mistakenly associates with the description, and this intended referent could be the basis of an intended statement that was true and constructively justified.

It is now possible to explain briefly why it is difficult to extend Donnellan's distinction beyond his 'favoured' examples to objects which cannot be 'given' to a speaker by pointing, such as ordinary visual objects at distant places or times, and theoretical or abstract objects. Extending the distinction in this way would be equivalent to extending over the same range the distinctions between effective and non-effective reference and between constructive and non-constructive reasoning. We would have to begin by asking, 'What is it for a speaker to be able or unable to pick out whom or what he is talking about from a set of objects he cannot observe?' (Donnellan sometimes seems to imply even in his examples that a speaker using a description referentially must already have observed his object, but this is surely too strong and does not follow from the idea that the speaker should *be able to* pick out his object.) The speaker would have to be 'given' the objects by description (it would be actually impossible to do this for many of the objects by pointing or acquaintance), e.g., he might be asked to consider all the spies who have ever lived. There would be a good case for holding that he would effectively refer to some of these objects, e.g., if he used a detailed description of a spy's exact position in a certain government or a proper name associated with basic information in standard reference books. There would also be a good case for saying that he only referred to some of them non-effectively, e.g. if he used 'The shortest spy ever' or a name the reference of which was fixed with such a description. And there would be a whole series of intermediate cases, in which it would be natural to speak of degrees of effective reference and of constructivity. (At this point it would be relevant to recall the comment of Kripke (1977) that Donnellan's distinction is not 'exclusive or exhaustive' (p. 257).) We would move from one stage to another by 'analogy and extrapolation', asking whether the use of a certain kind of term had some degree of effective reference because the information associated with it was sufficiently like that associated with other uses of terms already accepted as having effective reference. (The quoted phrase is from the discussion in Quine (1981) of undecidable sentences (p. 35).) It would be an enormous job to describe in detail why the use by speakers of certain terms for different kinds of objects enabled them to decide to some extent which particular object they were talking about and to reason about it constructively in some sense.

But at that point we would understand constructivity in natural language, and would be able to explain fully how the relation between it and Donnellan's distinction answers the question of the semantic significance of the distinction (much discussed by Kripke (1977) and others). An extreme constructivist would no doubt claim that the distinction would be highly significant semantically as marking the difference between legitimate and illegitimate uses of singular terms; he would say that the only legitimate use of a term was referential. (He would probably try to restrict himself to a Dummett-style effective language, in which the *meanings* of *all* the terms were given by effective methods.) And a classical semanticist would probably predict that the distinction would have only epistemological interest. But it can be shown that there is in turn a relation between constructivity and the propositional attitudes. Terms used with effective reference are 'suitable' for expressing *de re* and constructive attitudes, while terms used with non-effective reference are only 'suitable' for expressing *de dicto* and non-constructive attitudes. (See Over (1982).) Thus Donnellan's distinction will be at least indirectly relevant to the semantics of these attitudes.

*Sunderland Polytechnic,*
*Chester Road,*
*Sunderland SR1 3SD.*

© D. E. Over 1983

## REFERENCES

K. Donnellan, 'Reference and Definite Descriptions', *The Philosophical Review*, 1966, pp. 281–304.

—— 'Putting Humpty Dumpty Back Together Again', *The Philosophical Review*, 1968, pp. 203–15.

M. Dummett, *Frege: Philosophy of Language* (Duckworth, London, 1973).

S. Kripke, 'Speaker's Reference and Semantic Reference', *Midwest Studies in Philosophy*, Vol. II, 1977, pp. 255–76.

D. E. Over, 'Predicative and Constructive Knowledge', *Analysis*, 42.3, June 1982, pp. 140–6.

W. V. Quine, 'What Price Bivalence', in *Theories and Things* (Belknap Press, Cambridge, Mass., 1981).

---

# VESTIGIAL MODALITIES

*By* GEORGE I. MAVRODES

ACCORDING to Rabbi Harold Kushner, the Talmud says that if a woman is pregnant then neither she nor her husband may properly pray 'Let this child be a boy.' Why not? Kushner says the reason is that the sex of the baby was determined at conception, and now, the pregnancy having already begun, it is too late for God to do anything about it (Harold S. Kushner, *When Bad Things Happen to Good People*, Schocken Books 1981, pp. 115, 116). This reason is a special case of the more general claim that what has already happened has now a sort of fixity and necessity about it – it cannot, that is, now be affected, prevented, caused, and so on. I suspect that this claim is, in general, mistaken. Here, however, I want to explore just one of the ambiguities – a modal ambiguity – which makes the claim attractive.

This ambiguity can arise in theological contexts, but it has nothing *special* to do with such contexts or with God, religion, etc. And so my discussion here will have nothing special to do with these topics either. I will not, for example, discuss at all the claim that God is eternal, timeless, or outside of time. In fact, I will not mention God again.

Well, let's go. People often say something like

(1) The thing has already happened, so nothing can be done about it now.

Or, for a more specific example, consider

(2) She got married last week, so her father cannot do anything now to prevent her marriage.

A logician might analyse (2) as basically a conjunction of

(3) She got married last week.

and

(4) Her father cannot do anything now to prevent her marriage.

And perhaps he will add that the special conjunctive operator 'so' suggests that the second conjunct is derived from the first, supported by the first, or something of the sort.

I am inclined to think that (2) is indeed a conjunction, and that if (3) is true then (2) is true also. Of course, a conjunction does not generally follow from its first conjunct, so if (2) does follow in that way then its conjuncts must be related in some special way. But I think that (4) does not follow from (3). And I also think that (4) is not one of the conjuncts in (2). Or to put it more cautiously, (4), on its most natural interpretation, is not a conjunct of (2) and is not entailed by (2).

What is that interpretation? It is one which construes (4) as the modalized analogue of

(5) Her father will not do anything now to prevent her marriage.

It is the interpretation, that is, which takes (4) to ascribe some sort of necessity to (5) (or alternatively, to ascribe some impossibility to *He will do something now to prevent her marriage*). It is (4) under that interpretation which I take not to be a conjunct of, or entailed by, (2).

But can't we see just by looking at (2) that (4) is its second conjunct? No. But first let's look at another case which does not involve the relations of the future to the past. Someone trying to reconstruct some family history consults official records, and reports

(6) The County Clerk has no record of them, so they could not have been married here.

This appears to have the same form as (2), and its second conjunct appears to be

(7) They could not have been married here.

We might take that seriously and respond, 'Oh, that's too bad! I wonder why they could not have been married here. Do you suppose it was racial prejudice? Was the groom under-age according to the law here? You don't think the blood test...?' And so on. But the person who says (6) may reject all such suggestions. He may insist that there was no barrier to their marriage here. It was perfectly possible for them to be married here. No doubt they simply preferred some place else, and went there. And so on. And so the 'could not' in (6) seems to express an impossibility which is compatible with every possibility. What sort of impossibility is that?

There is a further and related puzzle. In (6) also the conjuncts seem to be related. How? One is inclined to say that one is derived somehow from the other, probably by some enthymematic inference. And what is the suppressed premise? Well, how about

(8) It is not possible that (both) the County Clerk has no record of them and they were married here.

It is plausible to suppose that (8) is true. The impossibility in it is, of course, not logical impossibility, but something like an impossibility growing out of laws, bureaucratic practices, etc. But there is some substance to such impossibilities. And it is also plausible to suppose that someone who did not accept (8) or something very much like it would not be likely to assert (6). Perhaps, then, (7) is derived from (8) and

(9) The County Clerk has no record of them.

And then we could say that (6) – the 'so' conjunction – is the idiomatic way of expressing or reporting that deduction.

The deduction of (7) from (8) and (9) has the following form.

(10) $\sim\Diamond(p \cdot q), p, \therefore \sim\Diamond q$.

Unfortunately, (10) is invalid for ordinary interpretations of the '$\Diamond$'. The impossibility in the first premise attaches to the *entire* conjunction, but perhaps not to any conjunct. In the conclusion it has been illicitly 'distributed' to one conjunct. But is there an interpretation of the '$\Diamond$' which makes (10) true? Well, maybe there is. Notice that if the '$\Diamond$' means nothing at all – if it is just a rhetorical flourish – then (10) is valid. Perhaps formal logicians never engage in such flourishes, but it would not be all that surprising if some uses of the modal terms in natural languages were of that sort. Notice also that if the '$\Diamond$' has its ordinary interpretation in its first occurrence in (10) but means nothing in the conclusion, then (10) again represents a valid inference. And that seems to me suggestive.

Here, then, is a conjecture: (6) is indeed a special conjunction, whose function is to express an inference from one of its conjuncts, (9), and a suppressed premise, (8). The proposition deduced, however, is not (7), but rather its unmodalized analogue.

(11) They were not married here.

And (11), I think, is the true second conjunct of (6).

The derivation of (11) from (8) and (9) is valid, thus accounting for our disinclination to reject (6) as illicit. And the fact that (11) is categorical, unmodalized, explains why (6) is taken to be compatible with there being no barrier to their marriage here.

Why, then, does a modal term appear in the last part of (6)? My guess is that it is the shadow of the corresponding genuine modality of (8). Its 'function' is to record and recall the modality of the suppressed premise, which does not otherwise appear in (6). Some biologists say that there are 'vestigial' organs which have no present function but which represent some of the evolutionary history of the animal which has them. In a similar way, I suggest that the 'could not' in (6) is vestigial. It has no genuine modal force *there*, and if we rely on it we are led astray. But it does represent the genuine modality of an inferential ancestor of one of the conjuncts in (6), an ancestor involved in the very inference which (6) expresses. Vestigial modalities are shadows without substance, but they are the shadows of a substance.

I think the same is true of (2). Its true second conjunct is

(12) Her father will not now do anything to prevent her marriage.

And this follows (validly) from (3) and

(13) It is not possible that (both) she got married last week and her father will now do something which prevents her marriage.

The function of (2) is to express this inference, and the modal term which appears in (2) is the vestigial representation of the genuine modality of the suppressed premise, (13).

If this conjecture is correct then (2), despite its appearance, is quite neutral with respect to the question of whether anything can be done now to prevent a past event, just as (6) is neutral about whether there was any barrier to having the wedding here. Perhaps there is some good argument which shows that nothing done now can have any bearing on the past. But I, at least, would be rather suspicious of any argument which appealed to our acceptance of propositions similar to (2).

*University of Michigan,*
*Ann Arbor, Michigan 48109, U.S.A.*

© GEORGE I. MAVRODES 1983

# LUCAS, LEWIS AND MECHANISM – ONE MORE TIME

*By* ANDREW D. IRVINE

MECHANISM can be understood as the thesis that minds can be explained as machines, or that for each mind there exists a mechanical model that will accurately portray it. J. R. Lucas has argued, in [6],[1] that certain definitive mathematically based objections can be made against the mechanist thesis. He has then gone on to defend his position, in [7] and [5], against criticisms made by Paul Benacerraf in [1], David Coder in [2], David Lewis in [3], and others. Nevertheless, the debate goes on. It is with some hesitancy that I enter into this debate. However, the most recent attack on Lucas's position, made by David Lewis in [4], fails. It would therefore be inappropriate to leave Lewis with the last word. It is the purpose of this note to indicate why Lewis's argument fails and perhaps to clarify Lucas's position by indicating the confusion which motivated the Lewis argument.

---

[1] The references are on p. 98.

In the context of this debate the term 'machine' refers to anything whose output, for fixed input, is capable of being encoded into a recursively enumerable set of Gödel numbers. Hence each machine can be considered to be a concrete instantiation of some axiomatizable formal system.

Lucas argues that it is not possible, even in principle, to have a mechanistic model of the mind. This is because, given certain plausible assumptions, there will exist for any given mechanistic model, M, of a mind, L, some well formed formula of the formal system which M is incapable of producing as being true (i.e., the formula is unprovable in the system). In contrast, since the formula can be proven true, or more accurately since L can produce a proof of the formula, no machine can be a complete or accurate model of the mind. For example, no formal system which is strong enough to include elementary arithmetic can produce as true an assertion of its own consistency. Lucas refutes mechanism since, for any M, Lucas can prove true a sentence, $\phi$M, which states that M's output is consistent, which he could not have done if he were M. (It is of course important to note that Lucas does not produce some sentence, $\phi$, which no consistent machine can produce.)

Lewis begins his attempt at refutation by introducing the following notation ([4], p. 374). $O_L$ is Lucas's potential arithmetical output (ignoring, for the sake of argument, any practical limits on Lucas's powers of computation), prior to the accusation that he is any particular machine. It is assumed that $O_L$ includes the axioms of elementary Peano Arithmetic, and that every sentence of $O_L$ is true. $O_L^M$ is Lucas's arithmetical output after he is accused of being some machine, M. We can understand $O_L^M$ to consist of $O_L$ plus the additional sentence $\phi$M, a Gödel sentence expressing the consistency of M's arithmetical output. In addition, let us introduce LO to denote Lucas's entire output (as opposed to simply his arithmetical output) prior to any accusation, and LO(M) to denote Lucas's entire output following the accusation that he is some machine, M.

Similarly, for any machine M, $O_M$ is M's potential arithmetical output when M is not accused of being any particular machine, and $O_M^N$ is M's arithmetical output when accused of being the machine N. $O_M$ and the various $O_M^N$'s may differ depending on the capabilities and complexity of M.

Lewis then argues that no contradiction arises if $O_L$ is identified with $O_M$, and $O_L^M$ with $O_M^M$, for some machine M. This follows since the sentence $\phi M$ must either express the consistency of $O_M$ or of $O_M^M$. In the first case $\phi M$ would not express the consistency of $O_M^M$ since $O_M^M$ contains some sentence (viz. $\phi M$), which is not contained in $O_M$, and thus $O_M^M$ and $O_M$ are distinct. Therefore, $\phi M$ can consistently be a part of an axiomatizable formal theory $\theta$ that has as its theorems all and only the sentences of arithmetic that are deducible in first order logic from $O_M^M$. Thus in this case the formal system of which M is an instantiation would be consistent, $\phi M$ would be true, and there would be no contradiction.

In the second case, where $\phi M$ expresses the consistency of $O_M^M$ there would again be no contradiction. This is because there is no contradiction in saying that because $\phi M$ belongs to $\theta$, $\theta$ must be inconsistent, and $\phi M$ false. This follows since there exists no proof that $O_M^M$ is consistent, or that $\phi M$ is true. The consistency of $O_M$ does not entail the consistency of $O_M^M$ since $O_M^M$ consists of $O_M$ plus the further sentence $\phi M$.

However, in order that Lewis's argument be sound, it must not just be the case that $O_L^M$ (or $O_M^M$) consist of $O_L$ (or $O_M$) plus the further Gödel sentence $\phi M$. It must also be the case that $\phi M$ is not included in $O_L$ (or $O_M$). Otherwise $O_L$ would be identical to $O_L^M$ (and $O_M = O_M^M$) and a contradiction would arise whenever $O_L$ is identified with $O_M$. Can this assumption be justified?

$O_L$ is Lucas's potential arithmetical output, that is the set of arithmetical sentences in first order logic that he is capable of proving before he is accused of being a particular machine, say M. However, the accusation that he is a machine, M, gives Lucas no new arithmetical information with which he can supplement $O_L$. The new information that Lucas gains (viz. that the mechanist is accusing him of being M), makes no difference to the arithmetical propositions which he is capable of proving, but instead determines the distinction between LO and LO(M). Thus if Lucas can prove the sentence $\phi M$ at all, it must be derivable from $O_L$. Since $O_L$ consists of Lucas's entire potential arithmetical output (i.e., any sentence deducible from $O_L$ is contained in $O_L$), $O_L$ will in fact contain $\phi M$. In order to understand this point it is helpful to recall that all Lucas needs in order to prove (or disprove) $\phi M$ is a specification of M, say in the form of a machine table. (At this stage he does not need to be aware of the mechanist's accusation that he is M.) $\phi M$ follows directly from properties of M, and no account need be taken of the mechanist's claim that Lucas is identical with M. $O_L$ will contain $\phi M$ just as it contains consistency results regarding classes of sentences other than $O_M$. Thus $O_L = O_L^M$, since $O_L^M$ consists of $O_L$ together with $\phi M$, and $\phi M$ is contained in $O_L$. Mistakenly Lewis assumed that $O_L$ was distinct from $O_L^M$, when really the only assumption justified by Lucas's position was that $LO \neq LO(M)$. Lewis's attempt refutation is therefore fallacious.

Why does Lewis assume that $O_L$ is distinct from $O_L^M$? This confusion arises from comments made by Lucas to the effect that the dialectical nature of the argument with the mechanist is necessary in order that he disprove the mechanist's claim ([4], pp. 373-4). Lewis has assumed (mistakenly) that Lucas meant he was unable to prove $\phi M$ until the mechanist accused him of being machine M. This is not the case since Lucas can prove $\phi M$ independently of the mechanist's claim. It is necessary that the mechanist specify M in his accusation only in order for Lucas to refute him, not in order that Lucas be able to prove $\phi M$. *What Lewis has not understood is that until the mechanist's claim is made, the fact that Lucas can prove $\phi M$ does not refute the mechanist thesis.* This is because Lucas has not proved some sentence $\phi$ which disproves the mechanist thesis for all machines. Rather, all that Lucas can do (and it seems to be enough), is disprove the mechanist thesis for any particular machine, M.

In summary, since it is the mechanist thesis that each mind has a particular mechanical model, Lucas must have a particular mechanical model – call it M. Since Lucas can prove (even prior to the mechanist's claim) $\phi M$, and M cannot, Lucas has shown mechanism to be false. (Thus there is a difference between LO and LO(M), because only after the accusation does Lucas know that mechanism is false, that the proof of $\phi M$ disproves the mechanist thesis, etc. However, none of these propositions, or their denials, would be contained in $O_L^M$.)

Lucas has claimed only that for any M he can prove a sentence $\phi M$ which M cannot prove. He does not claim that there exists a sentence which he can prove but that no machine can prove. The reason for this is that there exists an infinite number of machines that the mechanist might consider as possible mechanistic models for Lucas. It is interesting to note that if there existed only a finite number of consistent machines, say $M_1, M_2, \ldots M_n$, then Lucas could produce a single sentence that no consistent machine could produce (i.e., prove true). Such a sentence would simply be the conjunction of all the sentences $\phi M_i$, that is:

$$\phi M_1 \mathbin{\&} \phi M_2 \mathbin{\&} \ldots \mathbin{\&} \phi M_n$$

Since this is not the case, Lucas settles for the (apparently adequate) method of an argument schema which refutes each and every hypothesis of the mechanist.[1]

*University of Sydney,*
*Sydney, New South Wales,*
*Australia 2006*

© Andrew D. Irvine 1983

---

[1] I would like to express my thanks to the faculty and graduate students of the Department of Philosophy at the University of Western Ontario where an earlier draft of this paper was originally read, and also to a referee of *Analysis.*

## REFERENCES

[1] Benacerraf, Paul (1967), 'God, the Devil, and Gödel', *The Monist*, 51, pp. 9-32
[2] Coder, David (1969), 'Goedel's Theorem and Mechanism', *Philosophy*, 44, pp. 234-37
[3] Lewis, David (1969), 'Lucas Against Mechanism', *Philosophy*, 44, pp. 231-33
[4] Lewis, David (1979), 'Lucas Against Mechanism II', *Canadian Journal of Philosophy*, 9, pp. 373-76
[5] Lucas, J. R. (1970), 'Mechanism: A Rejoinder', *Philosophy*, 45, pp. 149-51
[6] Lucas, J. R. (1961), 'Minds, Machines and Gödel', *Philosophy*, 36, pp. 112-27
[7] Lucas, J. R. (1968), 'Satan Stultified: A Rejoinder to Paul Benarcerraf', *The Monist*, 52, pp. 145-58

# PERSONAL IDENTITY AND BODILY CONTINUITY: A FURTHER NOTE ON 'THE SELF AND THE FUTURE'

*By* HAROLD W. NOONAN

IN 'The Self and the Future' (*Philosophical Review*, LXXIX (1970). Reprinted in *Problems of the Self*, 1973. Subsequent page references will be to the reprint), Bernard Williams puts forward two arguments. One is an argument in favour of a mentalistic criterion of personal identity. The other is an argument in favour of a bodily continuity criterion of personal identity.

I have criticized the argument Williams gives for a mentalistic criterion of personal identity in a previous paper ('Williams on the Self and the Future', ANALYSIS 42.3, June 1982). In this paper I concentrate on the argument for a bodily continuity criterion, which also leaves me unconvinced. The person this argument is intended to persuade (hereafter 'the mentalist') is someone who accepts some form of a mentalistic criterion of personal identity but acknowledges familiar facts about the way people can suffer from amnesia, character change and illusory memories while retaining their identity.[1] Leaving aside the preliminary parts of Williams' argument for a moment, its heart is a challenge to the mentalist to draw a line somewhere in the following series of cases, i.e. to say at what point it becomes incorrect to say that the A-body person is still A after the change described:

(i) A is subjected to an operation which produces total amnesia,

---

[1] I hope my use of the term 'mentalist' will not be misleading. I use it to refer to anyone who holds the view that bodily identity is not a necessary condition of personal identity, i.e. that bodily interchange is conceivable. Clearly, any criterion of personal identity capable of justifying such a view must, if it is to be at all plausible, make reference to mental considerations, but it is not excluded, of course, from referring to physical considerations too.

(ii) amnesia is produced in A, and other interferences lead to certain changes in his character,

(iii) changes in his character are produced, and at the same time certain illusory memory beliefs are induced in him; these are of a quite fictitious kind and do not fit the life of any actual person,

(iv) the same as (iii), except that both the character traits and memories are designed to be appropriate to another actual person, B,

(v) the same as in (iv), except that the result is produced by putting the information into A from the brain of B, by a method which leaves B the same as he was before,

(vi) the same happens to A as in (v), but B is not left the same, since a similar operation is conducted in the reverse direction (*Problems of the Self*, pp. 55-6).

The mentalist must say that in (vi) the A-body-person is no longer A after the change described, since the B-body-person is. But, Williams asserts, he cannot deny that A survives the change described in (i). Hence he must draw the line somewhere in between, but this, Williams argues, he cannot do.

The best way to see the weakness of this argument is first to note its intended scope. The process of information transfer referred to in (v) and (vi) must be the fabulous process referred to in Williams' earlier argument for a mentalistic criterion of personal identity (*Problems of the Self*, p. 47), so the mentalist Williams is explicitly arguing against is one who accepts that such an information transfer constitutes a bodily interchange. But Williams refers at the beginning of his paper to Shoemaker's Brown/Brownson case and the criterion of personal identity it suggests according to which Brownson is Brown, namely: 'having the brain and in consequence the memory and character traits of', and he clearly intends his argument to be a challenge to this suggestion too (*Problems of the Self*, p. 47).[2] So let us consider how it might apply.

Of course, we now need to consider a somewhat different series of cases: (i)-(iv) can remain the same, but (v) and (vi), in which reference is made to information-transfer, need to be replaced by (v′) and (vi′) if the argument is to be relevant to Shoemaker's position:

(v′) the same as in (iv), except that the result is produced by dividing B's brain and putting one half into A's skull (A's brain having been removed) leaving the other half in B's skull,

---

[2] Shoemaker himself does not, in *Self-Knowledge and Self-Identity*, endorse the judgement that Brownson is Brown. But he does so in 'Persons and Their Pasts', *American Philosophical Quarterly*, 1971, where he also formulates a sufficient condition of personal identity, namely: 'having a total mental state M-connected by a non-branching causal chain to a previous total mental state of', whose satisfaction is entailed by the satisfaction of the just stated criterion. If Williams' argument refutes the suggestion in the text it therefore also anticipatorily refutes Shoemaker's position in 'Persons and Their Pasts'; this explains the references in the next paragraph and following to Shoemaker.

(vi′) the same happens to A as in (v′), except that the result is produced by transferring B's entire brain into A's skull and conducting a similar operation in the reverse direction.

Now Williams' challenge to the mentalist is to draw a line in this series, and thereby to show that he can reconcile his philosophical convictions with familiar common-sense facts about the way people can suffer from amnesia, character change and illusory memories while retaining their identity. But it seems easy enough for a mentalist of Shoemaker's persuasion to do so. The line, he will say, must be drawn between (iv) and (v′). Up to (iv) the changes are consistent with A's survival as the A-body-person, but with the change in (v′) identity of A and the A-body-person ceases. Williams objects to the proposal that in the series (i)–(vi) the line be drawn between (iv) and (v) on two grounds (*Problems of the Self*, pp. 56–7). First, that in the move from (iv) to (v) all we have added is that the 'memory' impressions which result from the change have, not only a model, as in (iv), but a model which is also their cause, and, he says, it is difficult to see why that, to A looking forward, could make the difference between expecting pain and not expecting pain (the A-body-person is to be tortured after the change in question is produced). Williams' second reason for denying that the line can be drawn between (iv) and (v) is that in (v) the A-body-person is certainly not B, for there we have an undisputed B in addition to the A-body-person, and certainly these two are not the same person. Hence, if the A-body-person is not A in (v), though he is A in (iv), this is not because he is B in (v), and who else, Williams is implicitly challenging the mentalist to explain, might he be?

Neither of these objections to drawing the line between (iv) and (v) applies to drawing the line between (iv) and (v′) in the revised series of cases. The difference between (iv) and (v′) is that in (v′) A's brain has been removed from his skull and replaced by one of B's brain halves, and in consequence the A-body-person is psychologically continuous with B: that is, the A-body-person is as he is psychologically because of the way B was in the past, and had B's history been different this would have been reflected in differences in the A-body-person's present psychological state. On the other hand, the A-body-person's present psychological state is not at all counter factually dependent on A's past history. It seems absurd to suggest that the outcome of this change, because it can be brought under the very general description 'the A-body-person has memory impressions which have not merely a model but a model which is also their cause', cannot rationally be regarded by A as involving his death (if he assumes his own brain to be destroyed) or, at least, his ceasing to have the body he previously had (if he is agnostic about the fate of his own brain). Certainly the mentalist can allow the common-sense facts about survival through amnesia, character change, and inducement of illusory memory impressions referred to in (i)–(iv) while maintaining that in (v′), as opposed to (iv), A ceases to be the occupant of the body which was previously his, which now comes to house a different person.[3] But then, who is now the A-body-person? This brings us to Williams' second objection to drawing the line between (iv) and (v), which again turns out to be no objection to drawing the line between (iv) and (v′) in the revised series of cases. This objection was to the effect that the A-body-person in (v) cannot be B, since there is an undisputed B in addition to the A-body-person in (v); so who *can* the latter be but A? In (v′), however, it is not so clear that we should say that there is an undisputed B in addition to the A-body-person. Rather, one might say, and a mentalist of Shoemaker's persuasion certainly would say, that what has happened to B is precisely what is referred to in the literature as 'fission'. Now how to describe fission is a problem to which there are several competing solutions, but none of them involves that just one of the fission products is the *undisputed* claimant to identity with the original person (on one view both of the fission products come into existence at the time of the fission so neither is identical with the original, on other views the two fission products are people who were once the same person as the original and the same person as one another). So if what happens to B in (v′) is regarded as fission then while it is certainly not true that the A-body-person is B, as Williams correctly says, nor is it true that the B-body-person is B; rather both are offshoots of B (however the notion of 'offshoots' is to be explicated) and, as for A, he is not in the picture at all.

Thus neither of Williams' two objections to drawing the line between (iv) and (v) carries much conviction when applied to the suggestion that the line be drawn between (iv) and (v′) in the revised series of cases and, what is the crucial point, nothing that he says makes it seem likely that a mentalist of Shoemaker's persuasion cannot retain his philosophical convictions while allowing for the common-sense facts about survival through amnesia, character change and inducement of illusory memories to which Williams refers.

---

[3] Amnesia may be partial or total, curable or incurable. Whether or not the amnesia referred to in (i) to (iv) is curable or not it cannot, despite what Williams says, literally be total: it cannot involve A's loss, not merely of the great part of his knowledge of his own past, but of *all* the knowledge and *all* intellectual capacities he had previously acquired – otherwise he would simply be unable to manifest the new character-traits referred to in (ii). So in (i) to (iv) the A-body's psychological state after the change will be to some extent at least counterfactually dependent on A's previous history and learning – in conformity, of course, with our everyday use of the term 'amnesia'. After the change in (v) or (v′), however, the A-body-person must be in exactly the same state as in (vi) or (vi′) if Williams' argument is not to be broken-backed. And in (vi) or (vi′) the A-body-person's psychological state after the change must be not at all counterfactually dependent on A's past history, otherwise the mentalist does not have to accept (vi) or (vi′) as a clear case of bodily interchange. Hence my confident assertion in the text that it is a *difference* between (iv) and (v′) that in the latter case the A-body-person's psychological state after the change is not all counterfactually dependent on A's past history.

If we turn to Williams' preliminary argument this conclusion is confirmed. He argues that if someone who had you in his power were to approach you and inform you that you were going to be tortured tomorrow, but that when the moment of torture came you would not remember any of the things you were now in a position to remember and would have quite a different set of impressions of your past, then you would be perfectly rational in fearing the torture (*Problems of the Self*, pp. 51–2). And so you would. But this is no argument against a mentalist of Shoemaker's persuasion since he can simply point out that the description given of what is to happen is general enough (once the question-begging repetition of 'you' is ignored) to cover both a situation in which the change described is brought about by a process of, say, brain-washing and hypnosis, which leaves your psychological state at least to some extent counterfactually dependent on your own past history, and a situation in which the change described is brought about by the removal of your brain and its replacement by another's. Since, according to the mentalist, *one* of these situations will involve your torture, it is entirely rational for you, convinced that what is described will happen, to feel fear, even if the other situation will *not* involve your torture – in just the way that, as Williams says, fear for yourself is rational if you are informed that 'one of you five – which one we've not decided yet – is to be shot' (*Problems of the Self*, pp. 58–9).

In sum, then, if Williams' argument is intended to be relevant to the mentalistic criterion of personal identity suggested by Shoemaker's Brown/Brownson case, as Williams gives us reason to believe at the outset of his paper, it is an entirely unconvincing objection to that criterion.

But, it may be said, all that this shows is that the scope of Williams' argument is more limited than he perhaps implies; it does not show that it is unsuccessful when used against the type of mentalistic criterion of personal identity to which he explicitly addresses himself.

However, once the ineffectiveness of Williams' argument against a Shoemaker type criterion of personal identity is seen, it does not take much thought to realize that it is equally ineffective even when regarded merely as an objection to the type of mentalistic criterion Williams explicitly discusses.

For the process of information transfer referred to in (v) and (vi) of Williams' series is meant to involve the total wiping clean of the brain to which the information is transferred, what Perry refers to in his review of Williams as a 'brain zap' ('Review of *Problems of the Self*', *Journal of Philosophy* 73, 1976);[4] thus in (v) the A-body-person is not merely in a state of amnesia with respect to A's life, from which he might emerge after a process of treatment or an accidental jolt on the head; his memories of A's life are gone as finally as they are in (v′), where A's brain is gone too. Given this it seems that Williams' objections to drawing the line between (iv) and (v) are as unsuccessful, so regarded, as they are when regarded as objections to drawing the line between (iv) and (v′). For the difference between (iv) and (v) is not merely that in (v) the A-body-person's memory impressions have a model which is also their cause, but that they have been induced by a very special causal process which ensures that the brain of A has been wiped clean of all the information it contained and that the A-body-person is psychologically identical with B. If the mentalist regards *this* difference between (iv) and (v) as a reason for drawing the line between them he would seem to be in no danger of having to deny common-sense facts about survival through amnesia (ordinarily so-called), character change and inducement of illusory memory impressions. Again, given that both the A-body-person and the B-body-person in (v) are psychologically identical with B, Williams' claim that we have an undisputed B in addition to the A-body-person after the change described seems as open to question as we saw it was in (v′). Again it seems that the mentalist can reasonably adopt the position that what has happened to B is fission and that both the A-body-person and the B-body-person are offshoots of B.

---

[4] It must involve such a brain zap, as I noted earlier, otherwise the mentalist would not have to accept (vi) as a clear case of bodily interchange.

If we turn to Williams' preliminary argument the same point is applicable: Williams' argument against the mentalistic criterion of personal identity he is explicitly attacking fails for just the same reason as it does if regarded as an argument against the criterion proposed by Shoemaker. For the description of what is going to happen which Williams imagines: 'You will be tortured, and will not remember anything about your present life, but will have a completely false set of memory impressions', is general enough to describe both a situation in which a brain zap has occurred and a situation in which amnesia as ordinarily understood has set in. Hence fear of torture will be appropriate, the mentalist can say, even if the result of a brain zap (without the information being transferred) is death; so the fact that fear would be appropriate in such a situation does not prove a brain zap without information transferral not to be death and does not show that there is anything wrong with the mentalistic criterion of personal identity under attack.

In sum, then, Williams' argument for a bodily continuity criterion of personal identity in 'The Self and the Future' is unconvincing. If, as I argued in the earlier article, the same is true of the argument he gives for a mentalistic criterion of personal identity, it follows that if there is any 'risk' involved in choosing between these two types of criterion, as Williams suggests, it is merely – so far as anything said by Williams in this paper goes to show – the uninteresting 'risk' of adopting a position on the basis of inadequate arguments.

Postscript: After writing the above I came across Robert Nozick's response to Williams' argument in his *Philosophical Explanations*

(Harvard, 1981). I think that the following brief comments on the differences between Nozick's reply to Williams' argument and my own may be illuminating.

Nozick's objection to Williams argument is that it presupposes the following (incorrect) principle: 'If x at time $t_1$ is the same individual as y at later time $t_2$, that can depend only on facts about x, y and the relationships between them. No fact about any other existing thing is relevant to deciding whether x at $t_1$ is (part of the same continuing individual as) y at $t_2$.' By contrast my own chief criticism of Williams could be put by saying that he *misapplies* this principle – which is a correct one – in arguing that the A-body-person must be A after the change in (v) if he is so in (iv), since the A-body-person's psychological state in (v) is not at all causally linked, as it is in (iv), to the A-body's past history.

Nozick implies at the outset of his discussion that he intends to explain how it can be rational for A, looking forward in (v), to regard the future A-body-person as himself and rational for him, looking forward in (vi), to regard the future B-body-person as himself. By apparent contrast my own response to Williams' argument entails that A ceases to exist in (v) and that the future A-body-person is merely one of B's offshoots.

However, despite Nozick's initial suggestion, his theory of personal identity as it is finally developed does in fact entail exactly the same description of case (v) as my own reply to Williams' argument, i.e., that A ceases to exist and the A-body comes to house one of B's offshoots (this is because, according to Nozick's theory: 'For a later stage y to be part of the same continuing object as an earlier stage x, not only must y be the closest continuer of x, also x must be the closest predecessor of y' (p. 42)). The real difference is that Nozick's theory entails that this is so even if it is the case *both* that the history of the A-body in (v) is exactly the same as that which it has in (iv) (with exactly the same (lack of) causal connections between later and earlier stages) *and* that in (iv) the A-body continues throughout to be the body of the person A. On my own view it is only possible to maintain that A ceases to exist in (v) but not in (iv) because of the difference between the history of the A-body in (iv) and its history in (v) previously pointed out.

*University of Birmingham,*
*Birmingham B15 2TT*

© HAROLD W. NOONAN 1983

# STEINER'S VANISHING POWERS

*By* JEREMY WALDRON

IF I have a right that you should do something, do I necessarily have a right to force you to do it? Hillel Steiner's latest contribution to this controversy (ANALYSIS 42.2, March 1982) contains an obvious mistake.

Steiner imagines an exchange between X and Y of the following three promises:

($P_1$): X promises Y that X will do A
($P_2$): Y promises X that Y will not enforce $P_1$
($P_3$): X promises Y that X will not enforce $P_2$.

Accordingly, X and Y fall under the following obligations:

($O_1$): X has an obligation to do A
($O_2$): Y has an obligation not to enforce $O_1$
($O_3$): X has an obligation not to enforce $O_2$.

Unfortunately for X, in Steiner's example, he has not managed to extract from Y any promise ($P_4$) not to enforce $P_3$. Thus, according to Steiner, Y has the following liberty (or privilege):

($L_4$): Y is at liberty (i.e., has no obligation not) to enforce $P_3$.

Suppose now that Y, sensing that X is about to default on $P_1$, sets about forcing X to do A. Is this act of enforcement permissible? Steiner's answer is that it seems to be *both* permissible *and* impermissible. It is impermissible inasmuch as it is a violation of $O_2$. But, Steiner argues, an act of enforcing $P_1$ may also be an act of enforcing $P_3$: one way to force X not to force Y not to force X to do A may be to force X to do A. So the act may be permissible inasmuch as it is an exercise of the liberty $L_4$. Thus, one and the same act is both permissible and impermissible. (And this, I guess, is supposed to show the absurdity of thinking that $P_1$ and $P_2$ are consistent, and hence the absurdity of thinking that we can drive a wedge between a right and a right to enforce that right.)

But it is simply not true that Y's act of forcing X to do A is permissible, on the premises we are given. In $P_2$, Y promised not to perform any acts of a certain type – call it type $T_2$ – and the act in question is of that type. It is true that he did *not* promise not to perform acts of another type, $T_4$, and maybe true too that all acts of type $T_2$ are acts of type $T_4$. But this leaves him, then, as far as we know, at liberty to perform all acts of type $T_4$ which are not also acts of type $T_2$. Clearly the act of forcing X to do A is not covered by that since it *is* an act of type $T_2$. Y's liberty, $L_4$, has therefore been rather sloppily described. It should be more carefully described as follows:

($L_4$*): Y is at liberty (i.e., has no obligation not) to perform any act of enforcing $P_3$ which is not also an act of enforcing $P_1$.

When the liberty is formulated in this way, it is transparent that forcing X to do A is not an exercise of it.

Steiner's mistake is to think that the liberty, $L_4$, has some sort of distinct existence (as though it were an inalienable liberty, or a constitutionally guaranteed immunity), so that it could be regarded as an open question whether $O_2$ overrides $L_4$ when they conflict. But $L_4$ has no distinct existence. It simply marks out the logical space that is left once $O_1$, $O_2$ and $O_3$ are given, and once we are assured that no further obligations are assumed. So $L_4$ cannot conflict with $O_2$, since it is *nothing but* a statement of what acts are *not* covered by $O_2$ (or any other obligation) so far as Y is concerned.

Thus Steiner has not shown the permissibility of the act deemed impermissible by $O_2$. Accordingly, he has failed to show that there is any absurdity or logical tension in the idea that one can have a right against someone without it being permissible to enforce it.

*University of Edinburgh*

© JEREMY WALDRON 1983

---

# IMPECCABILITY

*By* THOMAS V. MORRIS

MANY traditional theists hold that the propositions

(1) God is omnipotent

and

(2) God is sinless

are necessarily true *de dicto* and *de re*. They are necessary *de dicto* as partial explications of the concept of God. The properties of omnipotence and sinlessness are requisites for holding the divine office, for having the ontological status of being God. They are necessary *de re* by virtue of these properties being essential properties of the individual who in fact, and of necessity, holds this office. The individual who is God, Yahweh, is not such that he could have existed without being omnipotent and sinless. Indeed, his having these properties essentially is itself a condition of his being God. That is to say, the requisites for divinity having to do with power and morality can be spelled out more fully by the propositions

(3) God is essentially omnipotent

and

(4) God is essentially sinless

being propounded as necessary truths *de dicto.* So traditionally we have as conditions for deity essential omnipotence and essential sinlessness, or impeccability, which of course entail conditions (1) and (2).

It has been thought by some philosophers in recent years that (1) and (3) are logically inconsistent with the *de re* reading of (2), and thus of course with the acceptance of (4) as a necessary condition of godhood. If this were correct, the traditional concept of God incorporating (3) and (4) would be inconsistent, and of necessity there could exist no such divine being. In a well known article,[1] Nelson Pike has maintained that in order to avoid inconsistency at this point, the theist must abandon the impeccability condition and rest content with a requirement of merely actual, yet firm and dependable, sinlessness. According to Pike, the theist must hold further that the individual who is God could sin although he in fact never will do so. Thus Pike recommends rejecting (4) and the *de re* reading of (2) in order to maintain consistency with (1), and presumably with (3).

In his recent article, 'Omnipotence and Sin' (ANALYSIS 42.2, March 1982), W. R. Carter agrees with Pike that the consistency of the concept of God requires that the theist hold God's sinlessness not to be an essential property of his. However, Carter constructs an argument which he evidently takes to show that divine sinlessness could not be a contingent property. He concludes that Pike's move fails and suggests that we thus have good reason to question the consistency of the concept of God. I would like to explore the type of argument presented by Carter, point out a fatal flaw in the way he employs it, thereby defending Pike's move against his attack, reject that move for a different reason, and reiterate a basic point about the consistency of holding God to be impeccable as well as omnipotent.

Carter first points out that if sinlessness is a requirement for deity and if the individual who in fact holds the divine office is sinless only contingently, then it is possible not only for that individual to sin, but also and necessarily thereby to fail to occupy the divine office. And this is correct. The contingency of his godhood would follow from the contingency of his sinlessness and the

---

[1] Nelson Pike, 'Omnipotence and God's Ability to Sin', *American Philosophical Quarterly*, Vol. 6, No. 3 (July 1969), pp. 208–16; reprinted in *Divine Commands and Morality*, ed. by Paul Helm (Oxford, 1981), pp. 67–82.

*de dicto* necessity of (2). Carter attempts to express this point by saying that on Pike's position it would be possible for Yahweh to cease to be God through moral failing. In being contingently sinless, Yahweh could cease to be God through ceasing to be good, that is, through committing a morally reprehensible act or actualizing a state of affairs which would involve his contravention of some duty. He could, but never will.

Certainly, much that Pike says seems to support this understanding of what would follow from the *de dicto* necessity of (2) and the contingency of divine sinlessness. And it is against this that Carter directs his attack. He presents a simple but interesting argument intended to show that no divine being could possibly abdicate his position by lapsing into sin. As it stands, his argument contains a number of serious flaws. However, more rigorous instances of the type of reasoning can be produced to generate from the same basic premises the desired conclusion. I shall present here two versions, either of which alone would suffice to show that God cannot begin to sin.[2]

## ARGUMENT 1

Let us suppose, for *reductio*, that there is a possible world Ws in which an individual who in that world is God, say Yahweh, sins at some time *t* and thereby loses the status of deity, which he has enjoyed up until *t*. We must remember that omniscience, as well as omnipotence and sinlessness, is a standard traditional requisite for deity. Thus at a moment just before *t*, *t*—1, Yahweh must know that in Ws he will sin at *t*. None of the even remotely plausible arguments that omniscience does not entail foreknowledge of some kind or other would block this modest inference. And at that prior time, Yahweh either lacks or has the power to see to it that he does not sink into sin at *t*. This is merely the power to form an intention not to sin at *t* as well as the power to see to it that this intention is not thwarted. If he lacks the power, then Yahweh is not omnipotent and thus not God at *t*—1, contrary to our assumption. No adequate definition of divine omnipotence will allow a being to lack power to form his own intentions and still qualify as maximally powerful. Nor will a being count as omnipotent who cannot see to it that his own intended acts are performed. If on the other hand Yahweh has the power, then by stipulation of Ws, he chooses not to use it and so is not sinless, and thus not God, at *t*—1, again contrary to our assumption. God does nothing unintentionally. If he sins at *t* then he intends to sin at *t*. More precisely, if he sins at *t*, then he intends at *t* to sin at *t*. Further, God is never at a loss as to what to do under any circumstances. It follows from any adequate understanding of omniscience that for any of his conceivable future actions (indexed if need be to possible contingencies), God either intends to do it or intends not to do it. So if at $t-1$ Yahweh does not choose not to intend at $t$ to sin at $t$, that is, if he does not intend at $t-1$ not to intend at $t$ to sin at $t$, then he intends at $t-1$ to intend at $t$ to sin at $t$. But of course for God to intend to intend to $\phi$ is for him to intend to $\phi$. So at $t-1$ Yahweh intends to sin at $t$. Given that intentions to sin are just as much worthy of blame as the act or acts intended, an intention to sin is itself a sin. So at $t-1$ Yahweh sins and is thus not God, contrary to our assumption. And of course, strictly isomorphic arguments will apply to $t-2$ in reference to $t-1$, to $t-3$ in reference to $t-2$, and so on recursively to any prior $t-n$ in reference to $t-n+1$. Thus no possible world is describable in which a divine being is God for a while and then lapses into sin, thereby ceasing to be God.

---

[2] These arguments will assume that God is a temporal being. On the contrary assumption, the desired conclusion would follow much more directly.

## ARGUMENT 2

As in the first argument, we assume for *reductio* that there is a possible world Ws in which some individual, Yahweh, satisfies the requisites for deity up until a time $t$, at which time he sins, thereby ceasing to be God. We assume again that deity requires omniscience as well as omnipotence and sinlessness. At some time $t-1$ just prior to $t$, either Yahweh intends to sin at $t$, or he does not so intend. If he does, then in having such an intention he sins at $t-1$ and is thus not then God, contrary to our assumption. If on the contrary, he does not intend at $t-1$ to sin at $t$, it is either because he intends at $t-1$ not to sin at $t$, or because he has at $t-1$ no intention concerning whether he will sin at $t$. If the former is true, then by stipulation of Ws his intention is thwarted and he is thus not omnipotent, and so not God, at $t-1$, contrary to our assumption. Of course the intention had at $t-1$ not to sin at $t$ would be thwarted at $t$, and so it may seem that it is not until $t$ that Yahweh thus lacks omnipotence. However, there would be a power he lacks at $t-1$ – the power to see to it that his present intention is enacted at $t$. If on the other hand Yahweh at $t-1$ has no intention at all concerning whether he will sin at $t$, then at that prior time he fails to exemplify at least one requisite of deity and so fails then to be God. Either he lacks sinlessness, or he lacks omniscience, or both. In at least the case of an omniscient being such as God, being sinless entails intending never to sin. Thus, in order to be sinless at $t-1$, Yahweh would have to intend then never to sin, and thus not to sin at $t$. His lacking any such intention would entail his not being sinless and so not God at $t-1$, contrary to our initial assumption. In addition, it can be argued that if Yahweh were omniscient at $t-1$, then he would know at that time that he was about to sin at $t$. But God cannot know he is about to act in a certain way without intending so to act. This follows from the divine freedom (an implicate of omnipotence and omniscience). So if Yahweh did not intend at $t-1$ to sin at $t$, it would follow that he did not know at $t-1$ that he was about to commit this act. But in that case, he would not be omniscient and so not God at $t-1$, again contrary to our assumption. Once more our *reductio* is complete.

What is interesting about these two arguments is not their common conclusion, which any traditional theist could produce quite easily as an entailment of (4) or from an explication of divine immutability. But neither impeccability nor immutability is as relatively uncontroversial as any of the conditions for deity employed as premises in either line of reasoning. Therein lies their interest. Unfortunately, interesting as it might be, no such argument can serve Carter's purposes. The crucial flaw in his article lies in the way he tries to use this sort of argument to refute Pike's position. Carter thinks that if starting from the premises that God is omnipotent, omniscient, and in fact sinless now, we can establish that God could not cease to be sinless, then it will follow that he is sinless *essentially*, *contra* Pike. And of course this is false. If an individual can cease to have a property while yet continuing to exist, it is not an essential property of his. But from the fact that an individual cannot cease to have a property while continuing to exist it does not follow that it is one of his essential properties. The property of having been born on an island and the property of being a life-long unbeliever are both properties had by Bertrand Russell which he could not have ceased to have while yet remaining in existence; but regardless of how insular or incorrigible he might have been, they were not among his essential properties.

There may be a good *reductio* argument, say in the Anselmian mould, from an assumption of contingent divine sinlessness to an impeccability conclusion, but it will be quite different from the sort of argument used by Carter. In short, the kind of reasoning produced above and relied on by Carter cannot show that there is no possible world in which an individual such as Yahweh *fails* to be God through failing to be good; just that there is no world in which a divine being *ceases* to be God through ceasing to be good. And this could be true even if sinlessness were not an essential property of that individual's. So we must conclude that Carter's attempted refutation of the position on which God is sinless contingently fails.

What Carter overlooks is that there are quite a few other grounds on which to reject Pike's position, not least among them being the fact that it is a wholly unnecessary departure from orthodoxy. Peter Geach, another philosopher who believes there to be an inconsistency between omnipotence and impeccability, has chosen to denounce and eschew the former property, adverting to another conception of divine power labelled by him 'almightiness'. Yet other philosophers more recently have argued quite cogently that there is no real inconsistency to be avoided, that condition (4) and the *de re* necessity of (2) are perfectly compatible with conditions (1) and (3).[3] In either case, it is held that God's having maximal power over all things is logically consistent with his being sinless essentially. I believe this to be the correct position to hold.

The problem has been that if God is impeccable, or logically unable to sin, then it seems that there are possible actions for a being with his other properties which he is thus unable to perform, or possible states of affairs which he cannot in any sense actualize. And if this were true, then subject to all the required qualifications, it would be incompatible with God's being omnipotent. However, by drawing attention to a simple and independently plausible distinction, Joshua Hoffman[4] has suggested that any such acts or states of affairs we can describe or conceive are merely conceivable and not at all possible. In a less than Cartesian sense, the God who is impeccable is the ground of all possibility. Our ability to describe situations which would involve God's contravening some duty should just remind us of the distinction between conceivability and possibility. They do not coincide. And omnipotence ranges over only what is possible.

A careful explication of the notions of omnipotence and impeccability will allow both to be conditions of deity. On that basis, theists can accept conditions (3) and (4) as well as the *de dicto* readings of (1) and (2) without fear of inconsistency.

One final matter must be dealt with. In his argument against Pike, Carter wants to show that the theist both must and cannot consistently hold that God is impeccable. He suggests that there is a reason independent of the omnipotence problem that God cannot be essentially sinless. Referring to an article by Pavel Tichy, he says:

> Tichy argues that no individual is such that it is essentially non-malevolent. (103)

But what Tichy's argument is, Carter does not say. We can see why when we look at the passage referred to. In the process of trying to show that 'God' denotes an office or position rather than an individual, Tichy says:[5]

> If 'God' were simply the name of an individual, it would be a purely contingent matter whether God is benevolent or not; for *any* individual is conceivably malicious. But of course the notion of a malevolent God is absurd. If so, however, God cannot be an individual; God is bound to be rather something for an individual to be, and benevolence must be a part of what it takes for someone to be it.

---

[3] See for example Joshua Hoffman, 'Can God Do Evil?', *Southern Journal of Philosophy*, Vol. 17 (1979), pp. 213-20, and Jerome Gellman, 'Omnipotence and Immutability', *The New Scholasticism*, Vol. 51 (1977), pp. 21-37. See also Thomas P. Flint and Alfred J. Freddoso, 'Maximal Power', in *The Existence and Nature of God*, ed. by Alfred J. Freddoso (Notre Dame, forthcoming).

[4] Op. cit.

[5] Pavel Tichy, 'Existence and God', *The Journal of Philosophy*, vol. 76, No. 8 (August 1979), pp. 410-11.

As an argument that no individual is such that it is essentially non-malevolent, this does not amount to much. First, Tichy either just conflates conceivability and possibility or else assumes with no argument, and in the face of good arguments to the contrary, that the former entails the latter. And then he assumes that if a denial or contrary of a proposition is absurd, it follows that the proposition is necessarily true *de dicto* and not *de re.* Of course nothing of the sort follows at all. It is therefore clear that Tichy offers us here no independent reason to deny God's impeccability. The traditional theist thus seems to be on quite solid ground in affirming that God is impeccable, or essentially sinless, as well as being all-powerful. If there is any problem about impeccability, or the consistency of the traditional concept of God incorporating impeccability, it remains to be shown.[6]

*The University of Notre Dame,*
*Notre Dame,*
*Indiana 46556, U.S.A.*

© THOMAS V. MORRIS 1983

---

[6] I would like to thank Richard McClelland and Chris Menzel for their very helpful comments on an earlier draft and for convincing me to omit a complex elucidation of omnipotence, which I may advance elsewhere.

# Basil Blackwell

## Rationality and Relativism

**Edited by MARTIN HOLLIS and STEVEN LUKES**

The book takes its inspiration from the debate encapsulated in **Rationality** edited by Bryan Wilson (1970), but extends it into new fields and into current discussion. Martin Hollis, Steven Lukes, Robin Horton and Ernest Gellner contribute the developments in their own thought since the seminal articles included in that volume, and are joined by Ian Hacking, W. Newton-Smith, Charles Taylor, Jon Elster, Dan Sperber, Barry Barnes and David Bloor. Almost all the essays have been specially written for this book.
320 pages, hardback £16.00 (0 631 12773 9)
paperback £6.50 (0 631 13126 4)

## Modern French Marxism

**MICHAEL KELLY**

This is the first full-length study of French Marxist philosophy and social theory, from its origins to the present day. It provides a critical assessment of the way in which French Marxists have tried to understand human thought in relation to the world, and shows the historic circumstances which have shaped their efforts.
246 pages, £15.00 (0 631 13202 3)

## The Ethics of Environmental Concern

**ROBIN ATTFIELD**

The Ethics of Environmental Concern sets out to examine traditional attitudes to nature and to discover how far these need to be revised to cope with modern ecological problems. With its survey of moral traditions which are often misrepresented, and lively engagement of urgent contemporary problems, the book will interest all students of philosophy.
228 pages, c.£15.00 (0 631 13137 X)

## Choice, Welfare and Measurement

**AMARTYA SEN**

Amartya Sen, one of the most eminent economists of his generation, has made many outstanding contributions to economic theory. He has consistently tested the limitations of economic analysis, refusing to ignore social concerns such as justice and poverty. This book reflects the broad range of his pioneering essays on social choice, and his illuminating critiques of central economic concepts.
468 pages, £19.50 (0 631 12914 6)

**Basil Blackwell Publisher,** 108 Cowley Road, Oxford OX1 4JF

# The Philosophical Quarterly

Basil Blackwell Publisher is proud to announce publication, from January 1983, of *The Philosophical Quarterly*

*Volume 33 No. 131 April 1983*

| | |
|---|---|
| Dummett's Purge: Frege without Functions | *G.P. Baker & P.M.S. Hacker* |
| Acknowledgements of Equals: Hobbes' Ninth Law of Nature | *Joel Kidder* |
| J. R. Lucas, Kurt Gödel, and Fred Astaire | *David L. Boyer* |
| The Logical Paradoxes | *Graham Priest* |
| Natural Belief and Religious Belief in Hume's Philosophy | *Terence Penelhum* |
| DISCUSSIONS | |
| Penelhum on Hume | *Stewart R. Sutherland* |
| Peacocke on Wittgenstein | *Phillipa Foot* |
| CRITICAL NOTICES | |
| The Collected Papers of G.E.M. Anscombe | *Alan R. White* |

BOOK REVIEWS including a survey of recent work in the philosophy of education by JOHN HARRIS

A **Medical Ethics Special Issue** in 1983 will be devoted to philosophical discussion of first-order moral problems such as abortion, euthanasia and the treatment of mental illness.

**Frege Special Issue** 1984 Contributions should be submitted by 31 December 1983 to: The Executive Editor, The Philosophical Quarterly, The University of St Andrews, Scotland KY16 9AL.

**Subscription Rates 1983**

Individuals: £9.00 (UK) £9.00 (overseas) $22.00 (USA) $27.00 (Canada)
Institutions: £20.00 (UK) £20.00 (overseas) $45.00 (USA) $55.00 (Canada)
Orders, with remittance, or enquiries to Sue Dommett, at the address below.

**Basil Blackwell Publisher,** 108 Cowley Road, Oxford OX4 1JF

---

# MIND

A QUARTERLY REVIEW OF PHILOSOPHY
edited by D W Hamlyn
with the assistance of S D Guttenplan

**Vol XCII No 366 April 1983**

**Articles**

Sense, Assertion, Dummett and Denial: H. PRICE
Functionalism, Qualia and Causation: W. SEAGER
Art and Embodied Truth: B.H. BAXTER
The Status of Altruism: A. ROSS
Social Wholes and Parts: D.-H. RUBEN

**Discussions**

Dummett's Frege or Through a Looking-Glass Darkly: G.P. BAKER and P.M.S. HACKER
Decipherment, Translation, Interpretation: B.C. HURST
On Kripke's Puzzle: D.E. OVER
On a Property of a Perfect Being: M. LAZEROWITZ
The Anatomy of Thought: W.D. HART
Further on Actual Consequence Utilitarianism: M.G. SINGER

**Book Reviews**

**Books Received**

*Orders to:*
Journals Department,
Basil Blackwell Publisher Ltd.,
108 Cowley Road,
Oxford OX4 1JF, England

*Annual subscription:*
*Institutions:* £14.50 or $32.50 (Canada $37.75)
*Individuals:* £10.00 or $22.50 (Canada $26.25)
*Single issues:* £4.00 or $9.00 (Canada $10.50)

Published for the MIND ASSOCIATION by BASIL BLACKWELL PUBLISHER

## NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; T. R. Baldwin, Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. H. McDowell, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

PUBLISHER. ANALYSIS is published by Basil Blackwell Publisher Ltd, 108 Cowley Road, Oxford OX4 1JF, England. US Mailing Agent, Expediters of the Printed Word Ltd, 527 Madison Avenue, Suite 1217, New York, NY 10022, USA. Second Class postage paid at New York.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £11.00 (inland), £13.50 (overseas), $33.00 (USA), $39.50 (Canada); for individuals £7.00 (inland), £8.60 (overseas), $21.00 (USA), $25.00 (Canada). Each volume comprises four numbers, each of at least 48 pages, appearing within the academic year—in January, March, June and October. Orders should be sent to the publisher.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Mr Christopher Kirwan, Exeter College, Oxford OX1 3DP.

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about offprints. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by Whitstable Litho Ltd, Whitstable.

1 JUL 1983

ol. 43 No. 3 (New Series No. 199) June 19

# ANALYSIS

Edited by
CHRISTOPHER KIRWAN

## CONTENTS

| | |
|---|---|
| Report on ANALYSIS problem no. 19 | COLIN RADFORD |
| A benign regress | JOHN J. HALDANE |
| Identical truth-conditions | B. J. GARRETT |
| Popper, Tarski and relativism | RICHARD C. JENNINGS |
| Actuality and context dependence I | GRAEME FORBES |
| Actuality and context dependence II | MARTIN DAVIES |
| Doubting one's methods | ALAN R. WHITE |
| Keeping track of Nozick | CRISPIN WRIGHT |
| Reply to Hornsby on actions | E. J. LOWE |
| Events that are causings: a response to Lowe | JENNIFER HORNSBY |
| Actions, attempts and internal events | MICHAEL SMITH |
| Mr More on rigidity and identity | MICHAEL LISTON |
| Anomalous monism: reply to Smith | TED HONDERICH |
| More on counterpart theory | GRAEME FORBES |
| Is naturalism inescapable? | JEFFREY GORDON |
| Rationality and paradox: a reply to Conee | FRED KROON |

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003–2638

# Basil Blackwell

## The Philosophy of Leadership

CHRISTOPHER HODGKINSON

The behaviour of leaders has been widely studied, but strangely the philosophical aspects of the leadership function have been largely neglected. This book remedies this by developing a new theory of value for leadership, and is a logical development of Professor Hodgkinson's earlier work, Towards a Philosophy of Administration.
256 pages, c. £12.00 (0 631 13082 9)

## Kant's Political Philosophy

HOWARD L. WILLIAMS

Although recognised as an outstanding philosopher, Kant's political writings have been largely unappreciated. In this important new book Dr Williams demonstrates that Kant's political views are not only interesting in themselves but that they are essential to an understanding of his critical philosophy as well.
304 pages, £20.00 (0 631 13123 X)

## The Ethics of Environmental Concern

ROBIN ATTFIELD

The Ethics of Environmental Concern sets out to examine traditional attitudes to nature and to discover how far these need to be revised to cope with modern ecological problems. With its survey of moral traditions which are often misrepresented, and lively engagements of urgent contemporary problems, the book will interest all students of philosophy.
232 pages, £17.50 (0 631 13137 X)

## Marxism and Christianity

DENYS TURNER

Ever since Marx described religion as the 'opiate of the people', there has been open warfare between Marxists and Christians. Communism has been accused of being amoral, while Christianity has been condemned as ahistorical. Yet, to outsiders, both seem wedded to idealism with equal fervour. Denys Turner here examines the similarities and differences between the two competing 'world-views'.
268 pages, £17.50 (0 631 13118 3)

**Basil Blackwell Publisher,** 108 Cowley Road, Oxford OX4 1JF

# Basil Blackwell

## Metaphysics

Methods and Problems

GEORGE N. SCHLESINGER

What is metaphysics? What are the methods and techniques it employs? What kinds of problem fall within its scope? This lively introduction to metaphysics explores these questions in terms that students can understand and at the same time brings new insights to bear on perennial philosophical debates.

278 pages, hardback £18.50 (0 631 13124 8)
paperback £6.95 (0 631 13125 6)

## Bradley's Logic

ANTHONY MANSER

Once regarded as the greatest living English philosopher, Bradley's work is now largely neglected. In this valuable reassessment of **Principles of Logic**, Anthony Manser helps us to see how Bradley anticipated many of the problems explored by Frege and Russell, and demonstrates that it contains many of the questions which are still at the centre of philosophical discussion.

230 pages, £17.50 (0 631 13139 6)

## Plato: Hippias Major

Translated with Commentary and Essay by

PAUL WOODRUFF

Now accepted as genuine and as a crucial text in Plato's development of the theory of forms, **Hippias Major** here appears in a new translation by Paul Woodruff, accompanied by an extended essay discussing the dialogue as a philosophical whole, and examining its main themes.

230 pages, £12.50 (0 631 13091 8)

## Plato: Parmenides

Translated with Analysis and edited by

R. E. ALLEN

**Parmenides** is one of Plato's most obscure, even baffling dialogues, and there is considerable controversy surrounding its meaning. This new translation and analysis offers a remarkably clear and original discussion of the work, setting it in its philosophical context and taking full account of recent Plato scholarship.

384 pages, £25.00 (0 631 13121 3)

**Basil Blackwell Publisher,** 108 Cowley Road, Oxford OX4 1JF

ANALYSIS 43.3 JUNE 1983

# REPORT ON *ANALYSIS* 'PROBLEM' NO. 19

*By* COLIN RADFORD

If I am looking at myself in a mirror I am directly facing, do I see myself looking at myself? If so, do I also see myself looking at myself looking at myself – and so on?

EXCEPT for no answers at all, what the problem-setter fears most is a plethora of concurring answers showing that he is too easily puzzled. That didn't happen; there were twenty replies, much disagreement, and even the best answers struck me as persuasive rather than conclusive. Many were prolix where they might have been terse, but worse, even more were snappy where consideration was wanted. Concluding preliminaries with a customary remark that is nonetheless true: the overall standard was high, and almost every competitor may feel a little disappointment if he is not among the honourable mentions or in the money.

As Garrett wrote, it is 'invariably the case that when I look at myself directly in a mirror, I do see myself and I see myself looking at myself . . . .' We should need very powerful reasons to abandon either seemingly unexceptionable claim. Coersmeier abandoned both as a Berkeleyan, North also – and on further grounds similar to those suggested by the second question comprising the problem, viz. that their acceptance leads to an endless regress which is vicious and even less acceptable than their denial! But the only non-paradoxical answer to the first question is 'In normal circumstances, yes.' What about the second?

The most popular answer was that when I see myself looking at myself I *do* see myself looking at myself, looking at myself, etc., but this regress, though infinite, is non-vicious for the more complex statements are logically equivalent to the statement that I see myself looking at myself.

Hetherington denied the equivalence, arguing that even if 'He is looking at himself' entails 'He is looking at himself looking at himself', 'He sees that he is looking at himself' would not entail 'He sees that he is looking at himself looking at himself' because he might not 'notionally see', i.e., realize, this. (Rieber, McCormick, and Bedau-with-Monroe made related points.) However, Hetherington allows that 'Being the sort of person he is', and that is probably a 'Reflective Narcissist', 'R.N. probably *will* indulge in many more seeings-that while he and the mirror are facing each other.' But if the perceiver is reflective, won't he see that 'He is looking at himself' *does* entail 'He is looking at himself, looking at himself, etc.'?

(In which case, won't R.N. – and Hetherington – collapse into the position described in the most popular answer?) Hetherington doesn't say, and Jones argues that 'Just because the description' (of 'myself' as 'myself seeing an image of myself') 'does not strike me, does not mean that it is not necessary' and we must accept the resulting regress. Jones comforts us with his concluding remark, 'This is not really so bad: the T.V. camera which films its own monitor screen produces a similar regress which is perfectly natural, and not at all puzzling.' But *is* this regress similar?

Others denied that 'I see myself looking at myself' entails or is indeed equivalent to 'I see myself looking at myself, looking at myself, etc.' Some of them maintained that therefore the regress could not get going because, as Benado wrote, 'There is no further fact' beyond my seeing myself confronting myself in a mirror 'which via my perceptual and intensional states (e.g., knowledge) could affect my visual experience to generate it.' Others did allow such further facts and therefore maintained that a regress could get going but would soon break down. 'The absence of a theoretical upper limit does not guarantee an infinite perception in practice', wrote Sorensen, for 'Although the level of discrimination varies from person to person, I am sure it falls short of, say, ten iterated . . . clauses.'

Daley and Gonzales both argued that because reflections take time there can be no infinite extensions of looking at oneself. But as the problematic reflexivity is mental and conceptual, rather than physical – there is only one mirror – this argument is bad.

I must mention Yagisawa who gave a complex, impressive analysis of 'A sees A', which suggested an answer to the problem, but he did not give one.

What then is the correct answer to the second question? Consider a case in which two persons look at and see each other and A sees B looking at A, and B sees A looking at B. There seems little temptation to add that, therefore, A sees B looking at A, looking at B, etc., (or vice versa) although this clearly follows. It seems clear in this case that the repetitive statements, however extended, add nothing to the simpler ones from which they follow. Here the regress is patent, patently non-vicious, and quite unseductive, I think.

Consider next a case in which A is looking at himself with the aid of two mirrors, so that he sees the side of his head. Then A sees himself looking at himself, looking at himself, etc.[1] The temptation to produce the endlessly regressive statement again seems much less than in the puzzle case, as does the temptation to find anything mysterious in it.

I suspect then that these temptations are provided in the case in which the looker looks into (the reflections of) his own eyes by the feeling that looks are endlessly bouncing between him and what he looks at, viz. his own eyes, or, rather, their reflection. His eyes, and their reflections, are 'seen' as, i.e. seem to be analogous to, a pair of parallel mirrors. And since parallel mirrors would reflect each other endlessly, if only they and light were not physical phenomena, and since what we are dealing with is a mental and conceptual reflexivity, . . . . The confusion produces a substantive, and so an endless, vicious, puzzling, charming regress.

---

[1] Although we might not want to say that he literally sees *that* he is looking at himself, because he does not see his eyes – although he probably will notionally see, i.e., realize, that he is doing so.

The winner and recipient of £45 is Haldane, who not only asserts that the endlessly iterated statement follows from 'I see myself looking at myself' but does more than any other competitor to show that it does and why the regress is non-vicious. Garrett, whose answer is altogether more concise and tighter than Haldane's, is too short at this crucial point and must be content with £30 and second place. The dissenting Hetherington, having gone well over the word-limit, can win nothing.

*Keynes College, The University,*
*Canterbury CT2 7NP*

# A BENIGN REGRESS

## *By* JOHN J. HALDANE

IT may be profitable to consider the problem in stages.

### (1) DO I SEE MYSELF IN THE MIRROR?

Whatever is the causal story to be told about the involvement of the glass, there is no good reason to deny that what I see in the mirror is myself. Additional conditions must be satisfied beyond my presence before the glass for me to see myself by means of it; but the causal complexity of the case is matched in other circumstances, e.g. when something is seen lying beneath water. A reflection of an object, or its appearance through various media, are not proxies for the thing itself. They are simply ways in which it is presented – what one might choose to call 'object-environment composites'.

### (2) DO I SEE MYSELF LOOKING AT MYSELF?

Standardly, episodes of awareness can be analysed in terms of two components: the object of attention, and the means by which it is attended to, viz. the mental act. Epistemological realists, such as Aquinas and later Reid, have insisted on this distinction in their attacks upon Representative theories of cognition, and accused their opponents of confusing the two elements. In some circumstances, however, one's own cognitive acts can themselves be objects of attention. Such is the reflexive power of consciousness that one can think of x, and also attend to this x-intending thought; and a special case of this possibility is that in which x is the thinker. If I can think of myself (presently) thinking of myself, what then of perceptual awareness?

One may look at a scene and also be aware of the phenomenal qualities of one's perception. But this is neither a case of seeing oneself, nor of seeing oneself seeing. Indeed, in the normal way of things one cannot see oneself perceiving; and the explanation of this impossibility is that usually one cannot look at oneself at all. Unlike the thinker who is within range of his own thoughts, the perceiver is not a part of his visual field. This limitation is surmountable, however, and is overcome in situations involving reflective surfaces.

A subject, therefore, has two routes of access to his perceptual acts: an internal one revealing qualia; and an external one, provided in this case by the mirror, which brings himself within sight, and thereby makes possible his self-intending gaze. In this way I may see myself looking at myself.

## (3) DO I SEE MYSELF LOOKING AT MYSELF LOOKING AT MYSELF, ETC?

If (a) I see myself looking at myself, then (b) I see my looking at myself, and (c) I see myself. The second entailment has a single direction, but the first goes both ways, for (a) and (b) are logically equivalent. Further, it follows from (a) that I see myself looking at myself looking at myself, etc. This regress, however, is purely logical and is quite benign.

Just as the regress of entailments from an event's being 'present' to its being 'present in the present' etc, does not require an infinity of times; neither does the truth of the entailed 'looking' sentences imply an infinity of acts and objects. Rather these iterations logically reduplicate those items actually involved: my seeing, and its object – myself looking at myself. The regress amounts to no more than a logical elaboration of this reflexive activity, by repetition of the fact of the act's presence within its object. Every member of the generated series is logically equivalent to every other, but there is no non-logical relation of dependence of less distant members upon more distant ones; and hence no element of regressive viciousness.

*Birkbeck College,*
*Malet St., London WC1E 7HX*

© JOHN J. HALDANE 1983

# IDENTICAL TRUTH-CONDITIONS

*By* B. J. GARRETT

A NATURAL first thought concerning the two questions is that they are both to be answered in the negative since when I look in a mirror I am directly facing I do not, strictly speaking, see myself, rather I see a reflection of myself. And since reflections do not have intentional properties (in this case, the property of looking) no infinite regress threatens – it is blocked at the first stage.

However, this response overlooks an ambiguity in sentences of the form 'I see a reflection of myself looking at X'. If we take 'looking at X' to be predicated of my reflection, i.e. effectively to treat such sentences as equivalent to sentences of the form 'I see my reflection looking at X', then they do indeed involve a category mistake (reflections cannot look). But if, more naturally, we take 'looking at X' to be predicated of myself then such sentences are in perfectly good order – I see a reflection of myself and it is a reflection of myself-looking-at-X. And since the above questions can be reformulated in accordance with the latter interpretation (i.e. 'If I am looking at a reflection of myself, do I see a reflection of myself-looking-at-a-reflection-of-myself? If so . . . ?'), nothing is here gained by appeal to reflections.

An affirmative answer to the first question requires considerable qualification. To start with, it is quite possible for a person to look at X and yet not see X ('look' unlike 'see' is not an achievement word, in Ryle's sense). Moreover, even if we are guaranteed such a transition in this case it is still possible for a person to see X whilst X is $\phi$-ing and yet not see X $\phi$-ing, e.g. I might see John whilst John is looking at me, but not see John looking at me. However, despite the invalidity of the inferences from 'I look at X' to 'I see X' and from 'I see X' and 'X is $\phi$-ing' to 'I see X $\phi$-ing', it is nonetheless invariably the case that when I look at myself directly in a mirror, I do see myself and I see myself looking at myself, just as it is invariably the case that when I look at myself eating an apple, I see myself and I see myself eating an apple. In both cases the conclusions are equally legitimate and unambiguous (in the sense in which sentences such as 'I saw a man born in Jericho' are not unambiguous) though in the latter case there is no parallel to the second question, no threat of a regress.

Given this, is it true that if I see myself looking at myself then I also see myself looking at myself looking at myself and so on? And if it is true, is the regress which ensues virtuous or vicious? I think it is clear that there is a regress here but that it is entirely virtuous. For the truth-value of every statement in the regress is determined simultaneously and the truth-conditions of each statement in the regressive series are identical. In order to determine the truth of any statement in the series, one does not have first to determine the truth of the following statement in the series and thus the regress is not vicious.

To admit the existence of such a harmless regress is not (as perhaps the occurrence of 'also' in the second question might suggest) to admit the possibility of infinitely many distinct perceptions when I look at myself in the mirror. No such absurdities follow. The construction of such a regress, though syntactically and semantically legitimate, is of no philosophical importance.

*The University*
*St Andrews, Fife KY16 9AL*

© B. J. GARRETT 1983

---

# POPPER, TARSKI AND RELATIVISM

*By* RICHARD C. JENNINGS

PHILOSOPHERS of science are in general agreement that scientific realists must maintain the correspondence theory of truth. Scientific realists maintain that there is an objective world which science discovers – they maintain that the set of entities which make up the world, its ontology, is discovered in the process of doing science; that the world is not a fabrication of scientists made up to account for their observations. True scientific claims are true because in some sense they correspond to the objective world. The problem is to explain this correspondence – what is needed is a definition of truth in the sense of correspondence.

An obvious direction to turn, and a direction in which several philosophers of science have turned, is to the work of Alfred Tarski. Tarski's explicit purpose in his paper, 'The Semantic Conception of Truth' (*Philosophy and Phenomenological Research 4* (1943–4), pp. 341–75 – subsequent references to Tarski will be to this paper) is to give a more precise expression to the intuitions which give rise to the correspondence theory of truth. Karl Popper has turned to Tarski to find support for his brand of scientific realism, the idea of objective or absolute truth, and has used Tarski's work to support his view that science gets us closer to objective truth, and to lay the basis for his theory of verisimilitude. I will not discuss Popper's theory of verisimilitude in this paper; I will confine my attention to the foundational work which is based on Tarski. Popper's references to Tarski appear in various of his writings. One of the most developed discussions appears in his paper, 'Truth, Rationality, and the Growth of Scientific Knowledge' (*Conjectures and Refutations*, (London, 1963), pp. 215–50 – references to Popper are to this paper) and I will confine my discussion to this particular paper.

Because of the tendency of some philosophers to argue, as does Karl Popper, that Tarski's analysis of the correspondence theory of truth provides some foundation for scientific realism, it seems wise to look at the cogency of such arguments. Popper's arguments are among the most cogent and articulate and for this reason I choose to focus my attention on them. In the event, when one looks more closely at Tarski's own theory and arguments, it becomes evident that Tarski's theory, though it *does* justify using a correspondence theory of truth for science, *does not* justify its use in a way that affords any support for the scientific realist position, or, in the case of Popper, for the idea of objective or absolute truth. Moreover we discover that Tarski has specifically rebuked a critic who accused him of doing just this that Popper does. In what follows I will discuss Popper's idea of objective or absolute truth and show how he uses Tarski's work to support this idea. I will then discuss Tarski's work itself and show how it does not, and was not intended to, support such an idea as Popper's idea of absolute truth. Finally, I will conclude that Tarski's theory is compatible with a relativistic view of ontology.

## I. POPPER'S IDEA OF OBJECTIVE OR ABSOLUTE TRUTH

Popper maintains that:

> Thanks to Tarski's work, the idea of objective or absolute truth – that is truth as correspondence to the facts – appears to be accepted today with confidence by all who understand it. [Popper, pp. 224–5]

In the following passage we can see how Popper uses Tarski's analysis of the correspondence theory of truth to justify his claim that we can obtain objective or absolute truth:

> The highly intuitive character of Tarski's ideas seems to become more evident (as I have found in teaching) if we first decide explicitly to take 'truth' as a synonym for 'correspondence to the facts', and then (forgetting all about 'truth') *proceed to explain the idea of 'correspondence to the facts'*.
>
> Thus we shall first consider the following two formulations, each of which states very simply (in a metalanguage) under what conditions a certain assertion (of an object language) corresponds to the facts.
>
> (1) The statement, or the assertion, '*Snow is white*' corresponds to the facts if, and only if, snow is, indeed, white.
>
> (2) The statement, or the assertion, '*Grass is red*' corresponds to the facts if, and only if, grass is, indeed, red.
>
> These formulations (in which the word 'indeed' is only inserted for ease, and may be omitted) sound, of course, quite trivial. But it was left to Tarski to discover that, in spite of their apparent triviality, they contained the solution of the problem of explaining correspondence to the facts.
>
> The decisive point is Tarski's discovery that, in order to speak of correspondence to the facts, as do (1) and (2), we must use a metalanguage in

> which we can *speak about two things: statements; and the facts to which they refer.* [Popper, p. 224, Popper's italics]

Despite Popper's disclaimer, the force of the word 'indeed' in the above two formulations of truth conditions is very great. It suggests (and whether it is omitted or not does not matter once the suggestion is made) that through the metalanguage we have something like direct access to the facts. It suggests that in the metalanguage we can obtain the objective or absolute truth that Popper is seeking. And the suggestion is driven home by the final claim that the decisive point is Tarski's discovery that to analyse the correspondence theory of truth we must be able to speak of statements *and of the facts to which they refer.* But this suggestion is nothing that Tarski would tolerate. Let us briefly review Tarski's analysis.

## II. TARSKI'S SEMANTIC CONCEPTION OF TRUTH

It is certainly true that Tarski regards his semantic conception of truth as an analysis of the correspondence theory of truth. He states that the definition of truth should do justice to our classical conception of truth which, in modern philosophical terminology, can be expressed as, 'The truth of a sentence consists in its agreement with (or correspondence to) reality' [Tarski, p. 343]. But he regards this as insufficiently precise and clear to be considered as a definition of truth and proceeds to look for a more precise expression of this conception.

His first step is to provide a criterion of adequacy for any definition of truth. This criterion involves the equivalence scheme:

(T) X is true if, and only if, p.

From this scheme we obtain the now well known Tarskian equivalences such as '"Snow is white" is true if, and only if, snow is white.' These are obtained by replacing 'X' by the name of a sentence in the object-language (the language for which truth is being defined) and replacing 'p' by the sentence itself. He then states the criterion of adequacy as follows:

> ... we wish to use the term "*true*" in such a way that all equivalences of the form (T) can be asserted, and *we shall call a definition of truth "adequate" if all these equivalences follow from it.* [Tarski, p. 344, Tarski's italics]

The idea of correspondence is most fully developed in this criterion of adequacy for a definition of truth.

The actual definition of truth that Tarski arrives at involves the independently defined notion of satisfaction and is simply that, '... a sentence is true if it is satisfied by all objects, and false otherwise.' [Tarski, p. 353]. The point of this remark is simply to note that when philosophers such as Popper turn to Tarski in the process of developing or discussing correspondence theories of truth, it is to the criterion of adequacy of the definition of truth and the Tarskian equivalences that they turn, not to the definition itself.

Now let us consider whether the metalanguage, in which the equivalences of the form (T) are stated, gives us any further access to the facts than the object-language to which it is applied. We may be inclined, as Popper seems to be, to regard the metalanguage as providing a richer and in some sense more objective picture of the world. But for Tarski's theory such is not the case – the metalanguage is richer only in its linguistic and logical apparatus, not in its picture of the world:

> It is desirable for the meta-language not to contain any undefined terms except ... terms of the object-language; terms referring to the form of the expressions of the object-language, and used in building names for these expressions; and terms of logic. [Tarski, pp. 350–1]

The metalanguage must contain the object language as a part in order for the equivalences of the form (T) to be stated, but beyond this all it contains are terms for referring to the object-language and the terms of logic. In other words, the set of objects in the world to which the metalanguage can refer, its ontology, *is derived from* the object-language, and is no greater or less than that of the object-language. The only additional facts that the metalanguage gives us access to, beyond those of the object language, are certain linguistic and logical facts about the object-language itself.

## III. TARSKI'S CRITIC

In 'The Semantic Conception of Truth', Tarski replies to a critic who claimed that:

> ... due to the fact that a sentence like "snow is white" is taken to be semantically true if snow is *in fact* white (italics by the critic) – logic finds itself involved in a most uncritical realism. [Tarski, p. 361]

Tarski asserts that by using the words 'in fact' this critic creates the illusory impression that the semantic conception of truth is intended to establish the conditions under which we are warranted in asserting any given (empirical) sentence, when in fact the semantic definition of truth implies no more than that whenever a sentence such as 'snow is white' is asserted or rejected we must be ready to assert or reject the correlated sentence, 'the sentence "snow is white" is true.' Tarski's concern in dealing with this criticism is that there are those who may mistakenly reject the semantic conception of truth for various epistemological reasons. For Tarski,

> ... we may accept the semantic conception of truth without giving up any epistemological attitude we may have had; we may remain naive realists,

> critical realists or idealists, empiricists or metaphysicians – whatever we were before. The semantic conception is completely neutral toward all these issues. [Tarski, p. 362]

The moral of this story is that if the semantic conception of truth is such that we are not meant to reject it on the basis of such a position as, say, a subjective or a non-absolute theory of truth; then surely Tarski would also want to claim that we are not meant to accept such a position as the idea of objective or absolute truth on the basis of our acceptance of the semantic conception of truth. And yet this is just what Popper does. His use of the word 'indeed' discussed above has exactly the force of the words 'in fact' as used by Tarski's critic. These words suggest that the semantic conception of truth has something to offer us in establishing the conditions under which we are warranted in asserting empirical sentences; in particular they suggest that the metalanguage affords us some further access to the world that is not afforded by the object-language itself. But this is certainly not Tarski's intention, nor is it a justifiable interpretation of his theory. His theory, as I have shown, makes the access that the metalanguage has to the world entirely dependent on the object-language – the terms with which the metalanguage refers to the world are none other than those of the object-language.

## IV. RELATIVISM

The implications of this discussion are that despite its apparent utility in supporting the position of the scientific realist, Tarski's semantic conception of truth does not actually give such support. Nevertheless, it is an analysis of the correspondence theory of truth. But because the only non-logical non-linguistic terms that appear in the metalanguage are taken from, or have translations into, the object-language, the world to which the sentences of the object language are found to correspond is just the world of (presupposed by) that very object-language. The only kind of realism that can find support in Tarski's work is what Hilary Putnam calls internal realism (*Meaning and the Moral Sciences*, (London, 1978), pp. 123-40), and this is not sufficient for the scientific realist. The scientific realist is not satisfied with a realism according to which what is real in the world changes with changes in our scientific theories (or languages). Rather, he wants a form of realism which can justify the claim that science is discovering, or uncovering, the real world which exists independently of our theories. But this is just the realism that Tarski's theory does not provide – the metalanguage does not refer to any real world which exists independently of our theories, it only refers to the very same world to which our theories refer. Because the metalanguage is richer than the object-language *only* in its linguistic and logical apparatus, because its terms for referring to the world are taken directly from the object-language, the world referred to by the metalanguage is the very same world as the world referred to by the object-language. We should see the world that our metalanguage tells us about as a world our object-language creates. In this way we can see the semantic conception of truth as a form of relativism.

*Department of History*
*and Philosophy of Science,*
*Free School Lane, Cambridge CB2 3RH*

© RICHARD C. JENNINGS 1983

# ACTUALITY AND CONTEXT DEPENDENCE I

*By* GRAEME FORBES

## I

THE striking formal similarities between the model theories of tense logic and modal logic, deriving from their common treatment of special operators as quantifiers over entities of a certain kind, tends to suggest that these entities are on comparable metaphysical footings if tensed and modal discourse are on comparable semantic footings. So there is considerable philosophical interest in any argument to the effect that there are deep semantic differences between tense and modal operators. Such an argument, due originally to Gareth Evans, has recently been given an extremely clear presentation and elaboration by Martin Davies (see his *Meaning, Quantification, Necessity: Themes in Philosophical Logic*, Routledge and Kegan Paul 1981, pp. 201-9). As a result the present writer has a much improved grasp of what the argument is supposed to be; but this has only had the effect of increasing his scepticism about its claims.

When a speaker makes an utterance in a context C, the *content* of his utterance, that is, what he says, may be true absolutely ('TRUE') or false absolutely ('FALSE'). From these absolute evaluation predicates for contents of utterances, we can introduce relativized truth for the sentences uttered *via* the principle that the content of an utterance of a sentence s is TRUE iff s is *true at* the (index which constitutes the) context of utterance. Following Davies, let us also distinguish two kinds of operator O. If the content of a token of type 'Os' in a context C is determinable by treating the operator-type O as a function from contents to contents and applying it to the content which s has in C, then O is a sentence operator; so a sentence operator takes contents in a context to new contents in the same context. But if the content 'Os' has in C depends systematically on the content s has in other contexts, O is a context-shifting operator. Tense operators may be treated as either context-shifting or sentence operators, and what Evans and Davies try to show is that on either approach no analogous treatment of modal operators is available.

## II

Consider the temporally modified sentence

(1) Always (Whitlam is angry).

Davies writes (op. cit. p. 205): '... an utterance of "Whitlam is angry" ... in a certain temporal context, that is, at a certain time, is TRUE just in case Whitlam is angry at that time.' Hence, if an utterance of (1) in a context C* is to be TRUE, then for each context C, the content which an utterance of 'Whitlam is angry' has in C has to be TRUE (problems arising from the fact that 'Whitlam is angry' may not be uttered in each context can be overcome; the details are not germane here). This is the context-shifting treatment of 'Always', and so if an analogy between modal and tense operators is to hold, then

(2) Necessarily (Snow is white)

should be said to be TRUE iff the content of an utterance of 'Snow is white' in any context of an appropriate sort is also TRUE. Now it may seem clear that this is a possible account, since we can just take possible worlds to constitute the contexts, and we can agree that the content of 'Snow is white' varies from world to world just as that of 'Whitlam is angry' varies from time to time. However, Davies offers two objections to this move, (a) and (b) below.

> (a) Although 'Snow is white' has different truth-values at different worlds, if we regard this relativity to worlds as a kind of context-dependence, then, since content corresponds to what is said, 'we should have to say (absurdly) that we cannot know what has been said in an utterance of ["Snow is white"] unless one knows everything that is the case (that is, unless one knows which possible world is actual).' (p. 206)

Here Davies assumes that to know which world is actual is to know the totality of facts which actually obtain, but the analogous claim that to know which time is present is to know the totality of facts which presently obtain is clearly false, if such knowledge is required for grasp of the content of tensed sentences; so the assumption should be suspended awaiting the prospects of a parallel treatment of the two cases.

Davies' second objection to the context-shifting treatment of 'necessarily' is (loc. cit.)

> (b) ... we should have to say that if the world had been different in any respect then something different would have been said [by an utterance of 'Snow is white'] (that is, the very same utterance would have expressed a different complete thought).

No doubt such finely-discriminated contents are pre-theoretically peculiar. But Evans and Davies are happy to hold that utterances of 'Whitlam is angry' express different thoughts at different times, even when there is no difference whatsoever in the states of affairs which obtain at those times. So objection (b) seems merely to exhibit a prejudice rather than establish a conceptual difference.

## III

This particular skirmish is anyway rather incidental, for Davies provides good grounds to prefer the sentence operator treatment of tenses, in which we regard a token utterance of 'Whitlam is angry' as expressing a TRUTH-evaluable, context dependent content when it is the whole utterance, and as expressing a context-independent property of times if in the complete utterance it is modified by a temporal operator such as 'Always'. Thus (1) is TRUE in a context C iff all times in the past and future of the time of C, and this time too, are such that 'Whitlam is angry' is true at each. More generally, let us say that when 'Whitlam is angry' occurs modified by one or more temporal operators, it expresses a *timewise incomplete* content, or property of times, and the temporal operators preceding it map that content, operator by operator, into a more complex content, or property of times. Strictly speaking, these properties are all properties of the time of the context; for instance, 'Whitlam was angry' attributes to the time of C the property of being preceded by a time at which Whitlam is angry. A TRUTH-evaluable content is obtained by inserting the time of C into the content to eliminate its timewise incompleteness, which is a feature of primitive contents preserved by modification by tense operators. Since TRUTH-evaluable, i.e., complete, contents are supposed to be the contents of utterances, Evans concluded that utterances must be implicitly prefixed by an operator which he designated 'Now', whose function is to introduce the time of utterance into the content of the uttered expression.

On this approach, even 'Whitlam is angry' expresses an incomplete content, the complete thoughts we express by utterances of these words having the real form 'Now (Whitlam is angry)'. Davies holds that here there is again revealed 'a conceptual difference between modal and temporal cases... there is no reason to maintain that the expression ["Snow is white" as it occurs in (2)]... is anything other than a sentence which can be used, standing alone, to make a TRUTH-evaluable utterance.' Yet it is certainly customary to treat schematic letters in sentential modal logic as standing for expressions which are assigned properties of worlds and to treat modal operators as yielding more complex from less complex properties of this sort. So it is straightforward to extend the analogy with the temporal case to introduce context-dependence, by holding that each utterance is preceded by an implicit 'Actually', what follows this operator being an expression for a certain property, or incomplete content, the force of the utterance being that the property holds of the world which constitutes the context of utterance.[1]

Davies replies to this move that while TRUTH in the temporal case may be said to consist in truth with respect to a certain contextually determined time, the analogous claim about the modal case would be misleading: here there is just a basic notion of truth, and 'An utterance is TRUE just in case the sentence uttered is true. The actual world is that possible world with respect to which those sentences are true which are, in fact, true' (p. 209). But there is a problem with this 'basic' notion of truth. Davies mentions the 'familiar rubric': a sentence s is true just in case the world is as a person who used s to make an assertion would thereby assert it to be (p. 135). In this rubric, 'the world' presumably means the actual world, in which case, instead of being conceptually subsequent to some basic notion of truth, the idea of the actual world is involved in that notion. This restores the analogy with the temporal case: in each case, TRUTH can be thought of as truth with respect to a contextually determined index.

## IV

There is nothing in the semantic considerations which Davies advances which is incompatible with the possibilist doctrine that the meanings of modal sentences are given by their possible worlds translations, for these translations are simply quantified sentences, or sentences referring to a particular object, and can be regarded as made true or false by how things are, just as with sentences about material objects or numbers. The question at issue is rather whether our notion of context-dependence demands that we think of possible worlds as entering into the truth-conditions of utterances *via* some property of the utterance considered merely as an event, a property analogous to its spatio-temporal location. It is notable that when context-dependence is invoked, certain properties of objects are seen to be context-induced, the pastness of a time, the being local of a place, and so on. This suggests that one way of resolving the semantic issues raised by Evans and Davies is to tackle *directly* the metaphysical question of whether the property of actuality may be regarded as context-induced.

---

[1] A consequence of this analogy is that the content of a TRUE utterance is necessarily true, since if the actual world has a property, it has it at every world. In the same way, the content of a true utterance is eternally true. But if I say 'Grass is green', isn't what I say contingent? There is such a use of 'what I say'; but equally, if I say 'It's raining', there is a use in which what I say is now true, though it was false. See D. Lewis, 'Index, Context and Content', p. 97, in *Philosophy and Grammar*, eds Kanger and Ohman, Reidel 1981.

The thesis that actuality is context-induced, whose best known proponent is Lewis (see 'Anselm and Actuality', *Nous* 4 [1970] 175-88), should be sharply distinguished from the thesis that it is a contingent property of states of affairs or worlds. The latter is not disputed, just as it is not disputed that presentness is a temporary property of states of affairs or times. However, someone might hold that although contingent, actuality is an absolute (not induced) and primitive (not analysable) property of entities. Robert Adams has said that this thesis is confused (see 'Theories of Actuality', *Nous* 8 [1974] 211-31). He argues that since each world could have been actual – $(\forall w)\Diamond A(w)$ – then each world is actual at some world (itself, of course): $(\forall w)(Eu)A(w, u)$, by standard translation of '$\Diamond$', raising the degree of 'A' by one place for a world variable. From this he concludes that though the actual world is actually actual:

(3) Actually $(A(w^*))$

this means only that 'it is actual *in itself.* And every possible world is actual in itself' (p. 215). So there is no distinguishing $w^*$ from other worlds by a special property of actuality. But there is a fallacy here, for the property which (3) attributes to $w^*$ is not the property $(\lambda w)\mathit{Actual}(w,w)$, but rather the property $(\lambda w)\mathit{Actual}(w,w^*)$, which is the same property as is intensionally specified by '$(\lambda w)$Actually$(A(w))$', by standard translation of 'Actually'. This property is unique to $w^*$, and could be absolute and primitive.

However, we do not want to say the same about the present moment $t^*$ and local place $p^*$, so we must seek grounds which will not carry over to the modal case for denying that there are absolute and primitive properties $(\lambda t)\mathrm{Present}(t,t^*)$ and $(\lambda p)\mathrm{Local}(p,p^*)$. The significant difference between the modal case on the one hand and the temporal and spatial ones on the other is that we move through time and from place to place, but do not move through the other possible worlds in which we exist. Presentness and localness are properties whose changing application is tied to changes in our relations to the things to which they apply, while with modality there is no comparable phenomenon. And if movement is essential to context-dependence, a semantical implication would seem to be as follows. Suppose a speaker S is rooted to a spot all his life and he and other speakers use 'here' as a descriptive name whose reference is fixed by 'the place where S is located'. Then S's counterfactual 'If I had been located elsewhere, "here" would not have referred to here' is true, although 'here' is not an indexical. It therefore follows that it is not sufficient for the introduction of the apparatus of context-dependence that if things had been different, 'the actual world' would have picked out some world other than the actual one.

*Tulane University,*
*New Orleans,*
*Louisiana 70118, U.S.A.*

© GRAEME FORBES 1983

# ACTUALITY AND CONTEXT DEPENDENCE II

*By* MARTIN DAVIES

## I

IN his paper, 'Does Tense Logic Rest Upon a Mistake?' (to appear in a *Festschrift* for Donald Davidson, edited by Bruce Vermazen and Merrill B. Hintikka), Gareth Evans considered three possible semantic treatments of temporal modification. One he rejected as involving an incoherent conception of assertion. The other two he admitted as coherent; but in each case he argued that there are important disanalogies between temporal modification and modality. His conclusion was that tense logic should not be regarded as genuinely analogous to modal logic: we should not be 'blinded by the structural parallels between time and modality to their substantial differences'.

In *Meaning, Quantification, Necessity: Themes in Philosophical Logic* (Routledge and Kegan Paul, 1981; henceforth '*MQN*'), I stated the main points of Evans's argument. Graeme Forbes is still not convinced ('Actuality and Context Dependence I', ANALYSIS, this issue). In this note I shall try to make clearer what I take the argument to be.

Of the two semantic accounts that Evans admitted as coherent, one involves the claim that an expression, such as 'Whitlam is angry', which admits significant temporal modification is strictly speaking ambiguous. The other involves the claim that temporal modifiers are context shifting operators. On the ambiguity account, 'Whitlam is angry' is assigned quite different semantic values according as it occurs standing alone or modified by a temporal operator. As it occurs standing alone, the expression is a context dependent sentence, and relative to any context it is assigned a semantic value appropriate to a complete sentence. This value has to provide for the evaluation of an utterance of the sentence in that context as TRUE or FALSE once-for-all, where TRUTH is the primary dimension of assessment for utterances, the aim of sincere assertion. As it occurs modified by a temporal operator the expression is assigned (even relative to a context) a quite different semantic value, akin to that of a predicate of times. If we are 'semantically fastidious' then we should remove this ambiguity by distinguishing between what Evans called a *situation expression* (akin to a predicate of times) and the context dependent sentence which can be represented as the result of applying the closure operator 'Now' to the situation expression.

On the context shifting account, on the other hand, the expression 'Whitlam is angry' is a context dependent sentence in all its occurrences. But the semantic value assigned to a temporally modified sentence relative to a given context is not a function of the semantic value of the contained sentence relative to that same context; it depends upon the semantic value of the contained sentence relative to other contexts.

I shall consider the two accounts in turn (following Evans, and reversing the order of presentation in *MQN*).

## II

We can distinguish two lines of thought which lead to the idea that a 'semantically fastidious' representation of the context dependent sentence 'Whitlam is angry' will involve the closure operator 'Now'.

First, the situation expression 'Whitlam is angry' admits significant temporal modification because (even relative to a context) it expresses a condition that obtains at some times and not at others. But such a temporally relative condition cannot be the condition for the TRUTH of an utterance of 'Whitlam is angry' standing alone, since assessment for TRUTH is a once-for-all matter. To obtain an expression which expresses (relative to a context) a condition that can be a TRUTH condition we need to apply the closure operator to the situation expression.

Second, if our theory provides only for the temporally relative condition, then we are left asking the question 'What should [a sincere asserter] aim at?'. The answer, of course, is that as the language is actually used he should aim at producing an expression whose temporally relative condition obtains at the time of utterance. But in that case the expression uttered should be represented as involving reference to the time of utterance. Other conventions of use are possible, and if the convention were different then the representation would have to be correspondingly different.

It may seem that this is a distinction without a difference, but in fact the difference becomes clear when we turn to modality. If an expression is to admit significant modal modification it needs to be contingent. We can avoid potentially confusing complications if we choose an expression which is free of temporal or spatial relativity. So let us take as our example 'Snow is white', construed as expressing a condition which obtains only contingently but is not temporally or spatially relative.

Since 'Snow is white' already expresses a condition that can be the condition for absolute or once-for-all TRUTH of an utterance, we have nothing corresponding to the first line of thought. But it can seem plausible that there is, analogous to the second line of thought above, an argument for the idea that a 'semantically fastidious' representation of 'Snow is white' standing alone will involve the 'Actually' operator (which is formally analogous to the 'Now' operator).

The conventions of use of our language could have been such that an utterance of 'Snow is white' is TRUE iff the expressed condition *would have obtained (even) if kangaroos had no tails.* In that case the representation of the expression would have involved a counterfactual conditional. As the language is actually used, the utterance is TRUE iff the expressed condition obtains *as things actually are.* But in that case should not the expression uttered be represented as involving reference to the actual world? If this argument is a good one then we shall have to agree that no deep disanalogy between the temporal and modal cases has been demonstrated.

The main claim that Forbes makes is that a deep disanalogy cannot be demonstrated without appeal to metaphysical differences between times and possible worlds. I believe that Evans would have agreed. At one point he wrote: 'A sane view of possible worlds begins when we conceive of the actual world as the world in which all and only true propositions are true'. In *MQN*, I said essentially the same in reply to the argument set out in the last paragraph. (*MQN* p. 209. Christopher Peacocke drew the argument to my attention.) Forbes objects that the familiar rubric about truth reveals that notion as not conceptually prior to the notion of the actual world (Forbes p. 126). A short reply to that objection is that, as it occurs in the familiar rubric, the expression 'the world' does not refer to the actual world conceived as a member of a set of possible worlds. A more formal version of the rubric makes this clear: if *s* can be used to assert that *p*, then *s* is true iff *p*. (Cf. *MQN* p. 27.) In the next section, I shall set out a little more fully what I take the order of conceptual priority to be.

## III

My view about the location in the conceptual order of the notion of a possible world is similar to that expressed by Colin McGinn: 'if we are limning the true and ultimate structure of modal reality, we do better to leave the modal connectives primitive: paraphrasing them with possible worlds quantifiers misrepresents how things modally are'. ('Modal Reality', in Richard Healey (ed.), *Reduction, Time and Reality*, Cambridge University Press, 1981, p. 162.)

Given the ideas of a sentence with a truth condition and of assertion as aiming at truth, we can employ the notion of possibility (expressed by the modal operator) to arrive at the idea of a condition which does not obtain but might have obtained, and with it the idea of a sentence which is not true but might have been true. Given also the notion of conjunction, we can arrive at the idea that various conditions might have obtained, and corresponding sentences been true, *together*. Here we have the idea of a *possibility* (or a way things might have been, or even a way 'the world' might have been). The notion of a possible world is an idealisation of this notion of a possibility: a possible world is the limit of a process of refinement of possibilities. (See Lloyd Humberstone, 'From Worlds to Possibilities', *Journal of Philosophical Logic 10* (1981), and Graeme Forbes, 'Physicalism, Instrumentalism and the Semantics of Modal Logic', forthcoming in *Journal of Philosophical Logic.*)

Possibilities and possible worlds are, on this conception, abstract objects rather like sets of propositions. We can think of possibilities and possible worlds as regions and points in 'logical space', but there are clear ontological differences between these 'regions' and 'points' and intervals and moments of time, or regions and points in physical space. (See McGinn op cit. pp. 151–2.)

Thus we arrive at the idea of a collection *W* of possible worlds, and the idea of sentences being true or false relative to a member *w* of *W*. Amongst the possible worlds in *W*, one can be distinguished as the actual world: it is that world *w* such that the sentences that are true relative to *w* are precisely the sentences that are true (*simpliciter*). This is the conception of the actual world which, according to Evans, is needed for 'a sane view of possible worlds'.

If what I have just sketched is a sane view of possible worlds then, clearly, a proper conception of assertion and truth does not require the use of 'Actually' as a closure operator. For to say that in assertion one aims at truth relative to the actual world is to say no more about assertion than that one aims at truth.

It should be clear that, in offering these reflections, I am not disagreeing with Forbes's main claim. For whether or not they are taken in an eliminativist spirit, these reflections are metaphysical rather than semantic.

## IV

Let us turn to the question whether the temporal and modal cases can be seen as analogous if temporal modifiers are treated as context shifting operators. It might seem that the cases can be seen as analogous: 'we can just take possible worlds to constitute the contexts' (Forbes p. 124). But then, of course, we face the question whether we *can* think of worlds as contexts, and Forbes gives a reason for saying that we cannot: 'The significant difference ... is that we move through time and from place to place, but we do not move through the other possible worlds in which we exist' (p. 127). I think that Forbes is essentially right about this. I myself would try to highlight the significant difference by considering two facts about context dependent reference (e.g. the use of 'now' or 'here') in contrast to the use of 'actually'.

First, in context dependent reference the object referred to (e.g. a time or a place) is a member of an extensive class (of times or places). If we say that in a use of 'actually' there is reference to the world in the sense of 'I and all my surroundings', then we have a sharp contrast with this first fact. We avoid that contrast if we say that in a use of 'actually' there is reference to the actual world that is a member of the set *W*.

But, second, in context dependent reference there is an appropriate relation – an *en rapport* (causal) relation or a locational relation – between the token thought event (or token utterance) and the object referred to. On my conception, the members of *W* are abstract objects rather like sets of propositions, and we stand in no causal relations – and so no *en rapport* relations – to them. Nor can reference to a member of *W* be secured by the obtaining of a locational relation. For a token thought event or token utterance is located in countless different members of *W*.

Perhaps we can find something analogous, in the case of abstract objects, to an *en rapport* relation to a physical object. This is a difficult matter, but in the case of a set I think that the closest analogue is having exhaustive knowledge of the membership of the set. (See *MQN* pp. 110–11.) In the case of a possible world, the closest analogue would then seem to be having exhaustive knowledge of the conditions that obtain at the world. But that is knowledge that we do not have, even for the actual world. (Hence the remark at *MQN* p. 206, quoted by Forbes at (a) on p. 124.)

If a speaker uses 'now' or 'here' (or 'I' or 'you') in an utterance, then it is not sufficient for understanding of that utterance that an audience should merely know the meaning of the sentence uttered. He must be able to identify the time or place (or speaker or addressee) of the context of utterance; only then will he know what has been said. In central cases, this identification involves the audience thinking of the time or place (or person) under a demonstrative, rather than a merely descriptive, mode of presentation; and this is possible precisely because the audience stands in an appropriate *en rapport* or locational relation to the object. But one can hardly hold that understanding of an utterance involving 'actually' requires identification of the actual world from amongst the set *W* of possible worlds.

These reflections, like those of the previous section, are of a metaphysical, rather than a purely semantic, character. So again, they involve no disagreement with Forbes's main claim. Nor, I think, would Evans have disagreed. Concerning the application of a context shifting account to modality he wrote: 'To think in this way about possible worlds seems to commit one to an unacceptable form of modal realism'.

It seems then that on neither the ambiguity account nor the context shifting account of temporal modification is it genuinely analogous to modality.[1]

*Birkbeck College,*
*Malet Street,*
*London WC1E 7HX*

© Martin Davies 1983

---

[1] I am grateful to Graeme Forbes and Christopher Peacocke for advice on this note.

---

# DOUBTING ONE'S METHODS

*By* Alan R. White

A RECENT attempt in this journal (R. A. Sorensen, 'Knowing, Believing and Guessing', Analysis 42.4, October 1982, pp. 212-3) to support the thesis that knowing that p implies believing that p is based on the supposition that if one wrongly believes that one's method of arriving at an answer is unreliable, then one does not know the answer (Sorensen says 'Julia did not know the answers to the examination since she was mistaken about [Sorensen means that she had a wrong view about] the reliability of the method' and 'Since Jean falsely believes that his method is unreliable, Jean does not know the correct answer to the English history questions'). But this supposition is really a variant on the thesis at issue and so begs the question. Moreover, the supposition is as wrong, and for the same reasons, as the thesis.

Seeds of doubt about the reliability of his method may exist or be sown in the mind of one who uses a reliable method to arrive at an answer as easily, and in the same way, as seeds of doubt may exist or be sown about his knowledge of the answer in the mind of one who knows the answer. If someone is wrongly persuaded by a mathematician that his method of telling at a glance whether a large number, other than one ending in zero, is divisible by four, namely by checking whether the number represented by the last two digits is divisible by four, is mistaken, does he not know that, e.g. the number 1234567896 is divisible by four? If a physicist succeeds in making someone wrongly distrust the reliability of telescopes, does he not know that the distant object seen through his telescope is a building? If a sceptical philosopher's subtle arguments entice someone into falsely believing that sense perception is never a reliable method of discovering material objects, does he never know that material objects exist? There are also all sorts of reasons why I may wrongly doubt the reliability of a method which I am in fact currently and reliably using to gain knowledge.

Part of Mr Sorensen's reason – as of other philosophers – for supposing that mistaken doubts about one's methods show that one does not know the correct answer may be the assumption that one cannot know the correct answer unless one also knows that it is the correct answer. However, not only can this assumption be shown to be mistaken (see my *The Nature of Knowledge*, Rowman and Littlefield 1982) – it is a variant of the hoary thesis that in order to know something one must know that one knows it – but, more relevant here, doubt or disbelief in the reliability of someone's method of arriving at an answer, though perhaps a more plausible objection against his knowing that he knows it than against his knowing it, is not a valid objection even to the former. A man may not only know, but know that he knows, the answers to the above mathematical, physical and common sense questions even when he is beset by the doubts wrongly instilled in him by the evil geniuses there present. It is, of course, beside the point that someone who doubts the reliability of his methods would often, and rightly, not claim either to know the answer or to know that he knows it.

*The University*
*Hull HU6 7RX*

© ALAN R. WHITE 1983

# KEEPING TRACK OF NOZICK

By CRISPIN WRIGHT

MY purpose is to point out a lacuna in Robert Nozick's treatment of Scepticism in his recent *Philosophical Explanations* (Harvard University Press and Oxford 1981, pp. 167-288). It seems to me that, notwithstanding certain difficulties (see e.g. Raymond Martin, 'Tracking Nozick's Sceptic: a Better Method', ANALYSIS 43.1, January 1983, pp. 28-33), Nozick may well have succeeded in illuminating an aspect of our ordinary concept of knowledge but that, contrary to what he supposes, he has indicated no straightforward method of exploiting that illumination so as to obstruct the inferences characteristically drawn by classical forms of Scepticism.

## I

What is good about Nozick's proposal is easily illustrated by the following sort of case. Willow warblers and chiff-chaffs are so alike that only the expert can differentiate them. Suppose that a less expert Scottish birdwatcher sees what he rightly takes to be a willow warbler in his garden; he knows that, by and large, chiff-chaffs are only seldom found in Scotland; but, as it happens, a freakish local influx of chiff-chaffs has resulted in their temporarily substantially outnumbering the indigenous willow warbler population. Both 'justified true belief' and causal accounts are hard pressed to explain why this is not a case of knowledge: in particular, the amateur's justification for his correct identification essentially involves no false premise – it remains true that chiff-chaffs are, by and large, only seldom found in Scotland – and his true belief that he saw a willow warbler was directly 'non-deviantly' caused by the presence of one. Still he does not *know* what he saw: it was mere luck that he was right, – the bird could just as well have been a chiff-chaff *for all he knew.* Genuine knowledge cannot tolerate the involvement of that sort of good fortune.

The relevant sort of good fortune involves, evidently, a certain potential insensitivity on the believer's part to hypothetical variation in the fact whether or not P: *had* the bird been a chiff-chaff, he would still have believed that it was a willow warbler. Nozick's proposal, that a true belief that P is knowledge if and only if it is co-variant with the fact whether or not P across a range of relevant counterfactual states of affairs – if and only if, that is, it *tracks* the fact –, thus has a good deal of prima facie point. More specifically, X's true belief is knowledge if and only if

(a) Had it not been the case that P, X would not have believed that P; and
(b) Were it, in changed circumstances, still the case that P, X would still believe that P.

The proposal, as Nozick shows, needs various refinements; in particular, it will have to be insisted that relevant counterfacual circumstances not be ones in which X would have arrived at his belief whether or not P by a different *method*; and also that X would allow a certain priority to the results of the methods he actually uses over the results which other methods not meeting (a) and (b) would afford (cf. Nozick op cit. pp. 178–96). But these details have no bearing on our present concerns.

## II

Nozick thinks this aspect of the concept of knowledge can be used to reconcile the sort of inscrutable possibilities which classical forms of Scepticism so often exploit with many of our ordinary aspirations to knowledge. Suppose I am assailed with a sceptical doubt whether I have a hand, on the ground that I could, for all the difference it would make to my experience, be a brain in a vat whose conscious experience is wholly a function of a computer controlled input (=possibility SK). The sceptical thought is that since P (='I have a hand') is known by me to entail not-SK, my knowledge that P would put me in a position to know not-SK; hence, – since I cannot be in a position to know the latter, – it follows, contraposing, that I do not know that I have a hand (or anything else about my body or immediate physical environment). G. E. Moore sought, famously, to detach rather than to contrapose. But Nozick's thought is that both are wrong in the assumption that knowledge has to be transitive across (known) logical consequence. For *tracking* need not be: it may be the case that my true belief that P conforms with (a) and (b) – suitably refined –, that P is known by me to entail Q, and yet that my true belief that Q does not conform with (a) and (b).

It is natural to think that this cannot be right for the following reason. Suppose not-Q; i.e. that SK holds. Then I would have no hand but wouldn't know it; so how can it be the case that (a) is satisfied for X = myself and P = 'I have a hand'? Clearly, if my ability to track my possession, or lack, of a hand is to survive admission of the possibility of SK, then the range of counterfactual circumstances embraced by 'Had it not been the case that I have a hand ...' must fall short of SK and its kin. Why is Nozick entitled to presuppose that it does?

One argument for thinking that he is not is this. If the range of counterfactual circumstances embraced by 'Had it not been the case that I have a hand ...' is so restricted, then it ought to be true that:

(i) Had it not been the case that I have a hand, it would (still) not have been the case that SK.

But everyone can accept

(ii) Had it been the case that SK it would not have been the case that I have a hand.

Hence, by transitivity,

(iii) Had it been the case that SK, it would not have been the case that SK.

And (iii) is obviously unacceptable for anyone, who, like Nozick, wishes to concede to the Sceptic the point that SK is a real possibility.

But Nozick will not accept – presumably – that transitivity holds for counterfactual conditionals. And, prima facie, with good reason. The canonical form of denial of a counterfactual conditional is to claim that its antecedent *could* have held without the consequent holding. Clearly – prescinding from the case of logically true counterfactuals – *any* such denial would be correct if the range of relevant counterfactual circumstances is allowed to be wide enough. If someone asserts e.g.

(iv) Had he not been wearing a safety belt, he would have been severely injured,

he would be understood to be excluding the case where non-wearing of a safety belt was accompanied by a radical suspension of relevant laws of nature. The pointful assertion of counterfactuals rests upon a tacit understanding, shared by speaker and audience, of relevant *ceteris paribus* clauses. And unless the relevant *cetera* are the same for both premises, transitivity will be suspect. If I assert, for example,

(v) If there had been snow in the valley yesterday, I would have gone skiing,

you are not at liberty, putting that alongside the uncontentious

(vi) If an avalanche had then been taking place, there would have been snow in the valley yesterday,

to conclude

(vii) If an avalanche had been taking place yesterday, Wright would have gone skiing.

For it is at least likely that the range of counterfactual circumstances which I intended to embrace in the assertion of (v) excluded the case of avalanche, that I did not intend to be avowing my willingness to have gone skiing in suicidal circumstances.

What, though, determines the *extent* of the range of relevant counterfactual circumstances – the constitution, moving towards David Lewis' way of looking at these matters, of the class of possible worlds, such that if and only if P and not-Q hold in one of *these* worlds is 'If P had been true, Q would have been true', false? Lewis, as is familar, thinks that this is strictly a matter of the content of P: for any counterfactual, relevantly counter-exemplifying worlds have to be sought in a range accessible via *least disturbance* of the actual world compatible with the supposition that P is true. And, of course, on any account of this sort, transitivity must be unacceptable. Any case where P is a more 'distant' possibility than Q may be such that 'If P had been true, Q would have been true' and 'If Q had been true, R would have been true' may be acceptable, while 'If P had been true, R would have been true' is not: all that is required is that R holds in all the nearby worlds relevant to assessment of counterfactuals of the form 'If Q had been true . . .' in which Q holds; but fails to hold in the more distant worlds relevant to the assessment of counterfactuals of the form 'If P had been true ...' in which P (and Q) holds.[1]

If, and so far as I can see, only if this type of approach to the semantics of counterfactuals is broadly correct will Nozick be able, as his response to the sceptic demands, to endorse both (i) and (ii) without absurdity. But while he expresses the hope that his handling of subjunctive conditions will in no way prejudice a correct account of them, Nozick seems not to notice either his implicit purchase of the Lewisian line of goods; or its implausibility.

The salient point is that how 'far out', as it were, it is appropriate to search for relevant counter-examples to an asserted counterfactual looks to be determined not simply by its content but by the context in which it is asserted. *If* the matter were a function purely of content, determined à la Lewis, then if I asserted both (v) and (vi) in the same breath, you could have no cause to wonder whether I had indicated my willingness to undertake a suicidal skiing trip: I would have indicated no such thing – since, or so we may suppose, avalanches are 'further out' than normal snowfall in Swiss valleys in January. In fact, though, it is clear that in any normal conversational context, you would be bound to wonder – wouldn't you? – whether I had meant to hint that I had been depressed, or some such thing. If that is correct then the Lewisian account of the mechanism determining the relevant world-range for my utterance of (v) is wrong: rather, in order to determine that range, account must be taken of my simultaneous volunteering of (vi).

Again, consider

(viii) If Thatcher had been born and brought up a Russian, she would have been dedicated to the overthrow of the Western democracies; and
(ix) If Thatcher had been dedicated to the overthrow of Western democracies, she would have been a traitor to the UK.

Both might well be believable – relative to different ranges of possible worlds. But if someone were to assert both 'in the same breath', he would naturally be interpreted not as relying upon a failure of transitivity but as saying something extremely puzzling: what is the explanation of this puzzlingness, if not that the convention is that when a number of counterfactuals are in play in a single context, some *single* range of relevant worlds – perhaps that determined, à la Lewis, for the most remote of them – governs the assessment of them all?

The lacuna then, consists in Nozick's neglect – despite his generosity to himself with the extent of his lucubrations in general – adequately to research the semantics of conditionals. If the range of worlds relevant to assessment of an asserted counterfactual may vary from context to context, and if in particular it cannot be determined without consideration of what other counterfactuals are in play in the same context, then the acceptability of a counterfactual will become a context-relative matter. A consequent relativity will therefore infect the notion of tracking: there will, e.g., be no standpoint from which we can say, tout court, that (a) and (b) hold true for P = 'I have a hand', and not for P = 'I am not a brain in a vat'. All that it will be possible to say is that, in ordinary 'unphilosophical' contexts in which no possibilities like SK are in view, (a) and (b) will be considered acceptable for P = 'I have a hand'. But the Sceptic will insist that SK and its kin are possible circumstances in which (a) and (b) would not be acceptable for P = 'I have a hand'; and that my knowledge claim, failing the tracking condition in these circumstances, is therefore incorrect. How should we respond? If a context-relative view of counterfactuals is correct, we can scarcely deny – so long as we agree with Nozick that SK is a real possibility – that (i), in the context of the Sceptic's remarks, is unacceptable: it might have been the case that I had no hand precisely *because* SK held. All that Nozick would be able, lamely, to reply is that such is a very unusual context, and that ordinary uses of the word 'know' are conditioned by interpretations of the relevant counterfactuals of the form (a) and (b) which embrace much less remote possibilities (although it would be tendentious, in dialogue with the Sceptic, to describe such possibilities as remote). But no Sceptic denies that. What he wants to know is why the possibilities on which he chooses to concentrate do not betray the intended *point* of these knowledge claims. To call attention to our customary limited horizons by way of response to that query would be, in effect, to fall back on a 'paradigm-case' response after all: it would be ironic if, to borrow a jest of Simon Blackburn's, Nozick's shiny new clothes were to prove to conceal that particular boring old Emperor.[2]

---

[1] Cf. David Lewis, *Counterfactuals*, Blackwell, 1973, pp. 32–6; also Jonathan Bennett's critical study of Lewis in *Canadian Journal of Philosophy*, vol. 4, 1974, pp. 384–5.

* * *

In summary: Nozick seeks to explain how, consistently with the conceded impossibility of my knowing that I am not a brain in a vat, I might yet know that I have a hand. The explanation consists in the claim that knowledge, conceived as a form of tracking, will fail of transitivity across known logical consequence. This claim if true requires failure of transitivity of the subjunctive conditional. But Nozick provides no reason to accept the tradition in the literature which holds that there is indeed such a failure; indeed, he believes he has avoided controversy on such matters. I suggest that pursuit of the thoughts sketched above would show that tradition to be incorrect: subjunctive conditionals, like all conditionals, sustain transitivity in context. If that is so, all that remains of Nozick's way with the Sceptic is a version of the paradigm-case argument as sketched.

---

[2] Nozick, op cit., is (rightly) harsh with the Emperor on pp. 55–6 and in a footnote on p. 168.

Actually, Nozick is surely wrong to concede so much to the Sceptic. Is tracking *always* necessary for knowledge? There is no discussion of *On Certainty* in *Philosophical Explanations.*[3]

*Department of Logic and Metaphysics*
*The University,*
*St Andrews, Fife KY16 9AL*

© CRISPIN WRIGHT 1983

---

[3] My thanks to Christopher Bryant, Anthony Ellis and Bernard Mayo for helpful discussions.

---

# REPLY TO HORNSBY ON ACTIONS

*By* E. J. LOWE

JENNIFER HORNSBY concedes that it is a consequence of her theory of action that, where a human agent *A*'s action was his causing an event *x*, *A*'s causing *x* caused *x* ('Reply to Lowe on actions', ANALYSIS 42.3, June 1982). This commits her to saying, for example, that if John raised his arm (i.e., caused his arm to rise), then John's causing his arm to rise caused his arm to rise. I continue to find this consequence 'distinctly odd', notwithstanding Hornsby's bald assertion, in her reply, that 'So far as "*A*'s causing *x* caused *x*" is concerned, this is no odder than the plainly truistic "*e*'s cause caused *e*"'. For why should we accept in the case of human agents what is manifestly absurd in the case of inanimate agents? When I say that billiard ball *B* caused billiard ball *C* to move, i.e., caused the event of *C*'s moving, I certainly don't want to imply that *B's causing C to move* caused *C*'s moving. If I want to refer to an *event* which caused *C*'s moving, I may well refer to *B's moving*, but certainly not to *B*'s causing *C*'s moving (which is doubtfully an *event* at all).

Perhaps it will be objected here that the sense in which a billiard ball 'causes' another to move is quite different from the sense in which a human agent 'causes' his arm to rise. But I see no reason at all to suppose that 'cause' is thus ambiguous, and without a reason the supposition is extravagant; moreover, it is a supposition which is difficult to square with a naturalistic account of human beings. What no doubt *is* true is that if a human agent *A* caused his arm to rise, and we want to refer to the *events* which caused his arm's rising, we must be prepared to refer *at least once* to a very different sort of event from the sort of event we could refer to in the case of an inanimate agent such as a billiard ball. In the billiard ball example just cited, we could refer to *B's moving* as an event which caused *C*'s moving; and *B*'s moving is a purely physical event. Now, analogously in the arm-raising example, we may expect there to be events which we could refer to as having caused *A*'s arm to rise, and some of these events will again be purely physical, e.g., the contracting of *A*'s muscles: but we must also expect at least *one* such event, say *A*'s $\phi$-ing, to be a *mental* event of some sort (which is not, however, to deny that it may *also* be physical, only to deny that it is *purely* physical, like *B*'s moving or the contracting of *A*'s muscles). (What $\phi$-ing might be is a large issue I cannot go into here: some would say that it is 'willing', and I am not in principle opposed to this answer.) We can see, then, that what differentiates a human agent's causing something from an inanimate agent's causing something need not be supposed to reside in any difference in the sense of 'cause', but only in the fact that human agents, unlike inanimate agents, have minds and that mental events are causally efficacious. But if this is correct, there is no escaping the absurdity of saying such things as '*A*'s causing *x* caused *x*'. (For an admirable clarification of some of these matters, see Kent Bach, 'Actions are not Events', *Mind* LXXXIX, January 1980.)

*University of Durham,*
*50 Old Elvet, Durham DH1 3HN*

© E. J. LOWE 1983

---

## EVENTS THAT ARE CAUSINGS: A RESPONSE TO LOWE

*By* JENNIFER HORNSBY

IT would be strange if I were to say 'My writing this is caused by Lowe's causing my writing this'. This would be strange because (for one thing) there are so many more informative and more elegant sentences which convey a similar message and which I could use instead.

The uninformative and inelegant sentence is from the pattern of

(*) A's causing *x* caused *x*.

And Lowe and I agree that it is often possible to deduce an instance of (*) from premises that I should accept. But Lowe continues to find (*) 'distinctly odd' where 'A' stands for a human agent; and instances of (*) in which 'A' stands for something inanimate he now calls 'manifestly absurd'.

Lowe's original reason for deriving (*) himself was that he hoped to provide a *reductio*: he intimated that he had cast in doubt my assumption that actions are events (see ANALYSIS 41.3, June 1981, p. 129). Now that he has encouraged us to see the matter in a more general light (see above), I think he has made it clear that if we must reject the assumption that actions are events, then so also must we reject the claim that ball B's hitting ball C is an event.

The reader who shares Lowe's view of (*) may want to deny that B's hitting C is an event, or he may think that Lowe's reasoning was flawed. For my part, I am convinced that, however strange it might be to use them, many instances of (*) can be got from unexceptionable premises by truth-preserving steps.

*Corpus Christi College,*
*Oxford*

© JENNIFER HORNSBY 1983

---

# ACTIONS, ATTEMPTS AND INTERNAL EVENTS

*By* MICHAEL SMITH

OVER the past few years several people have argued that all tryings are internal events.[1] They claim to show that if, for instance, an agent tries to raise his arm, then whether he raises his arm or not, the rising of his arm is no part of his attempt. His attempt is rather an internal event which, if he does raise his arm, causes his arm to rise. The time has come to review such arguments. For, as Jennifer Hornsby has recently pointed out, if all attempts to act are internal events, and all successful attempts are identical with the actions the attempted performance of which was successful, then all actions are internal events. If such arguments can show that no attempt to raise an arm has arm-rising as a part, then no arm-raising has arm-rising as a part either. Her argument is valid, so those of us who think that her conclusion is false must reject one of her premises.[2] I reject the first. My assumption will be that the arguments given for the first premise have a common form. The contention I wish to defend here is that that form is fallacious.

---

[1] See for instance, Hugh McCann's 'Is Raising One's Arm a Basic Action?' In *Journal of Philosophy* May 1972; David Armstrong's 'Acting and Trying' in *Philosophical Papers* May 1973; Brian O'Shaughnessy's 'Trying (as the Mental "Pineal Gland")' in *Journal of Philosophy* July 1973; and, more recently, and more tentatively, Jennifer Hornsby's 'Arm Raising and Arm Rising' in *Philosophy* January 1980, and Chapter Three of *Actions* (Routledge and Kegan Paul, 1980).

[2] Amongst those who think that Hornsby's conclusion is false is E. J. Lowe: see his '"All Actions Occur Inside the Body"' in *Analysis* 41.3, June 1981. But Lowe's arguments against Hornsby are, I think, fallacious. See her 'Reply to Lowe on Actions' in *Analysis* 42.3, June 1982.

Why have people thought that all tryings are internal events? They accept an argument like the following. Suppose that an agent tries to raise his arm and raises his arm. Note that, had he been totally and unexpectedly paralysed just prior to trying to raise his arm, he might still have tried to raise his arm, though his arm would not have risen. In such a case his attempt would have been identical with some event, though, evidently, arm-rising would have been no part of it. In such a case his attempt could only have been an internal event. But now note that there is an event which occurs when an agent raises his arm, namely, his trying to raise his arm, and that the same event, the attempt, would have occurred had the agent been paralysed and the arm not risen. The latter attempt would have been an internal event, so, given its identity with the former, his trying to raise his arm when he raises his arm must be an internal event too. This argument can be generalized in the obvious way to show that all tryings are such internal events.

What is wrong with this argument? To begin with, one premise is ambiguous. It is claimed that when an agent tries to raise his arm and raises it 'the same event, the attempt, would have occurred had the agent been paralysed and the arm not risen'. Is it claimed that the same *type* or *token* event would have occurred had the agent been paralysed? It is uncontroversial that the event would have been of the same *type*, *viz.* an *attempted arm-raising*. But given this uncontroversial claim, and the additional fact that the attempt which would have occurred would have been an internal event, nothing follows about which *token* event the *actual* attempt is (and hence it doesn't follow that the actual attempt is an internal event). So, if the premise is read uncontroversially as 'same type', the argument is invalid.

It may be thought that if the premise is read as stating that the same *token* event, the attempt, would have occurred had the agent been paralysed, then the argument is valid. However, this thought is false if you take certain views on the cross-world identity conditions of events. For instance, if you hold that not all parts of events are essential parts, then the argument is invalid. In order to see this, consider an argument of the same form. Just as we supposed that there is an attempt to raise the arm and the arm rises, suppose that there is a party, and that towards the end of it, Monica has a last drink. The argument we are considering would have us reason as follows. Had Monica not had that last drink (the arm not risen), the same token event, the party (the attempt), would still have occurred. Monica's having that last drink (the arm's rising) would have been no part of the party (the attempt) which would have occurred. Conclusion: Monica's having that last drink (the arm's rising) is no part of the actual party (attempt). The conclusion doesn't follow if you think that the same token party would have occurred whether Monica had had that last drink or not: that is, if Monica's having that last drink is a non-essential part of the party which actually occurs. Thus, even if the crucial premise is read as stating that the same token attempt would have occurred had the agent been paralysed, given this view on the cross-world identity conditions of events the argument is of a form which is invalid.

Such a reply to the argument will appeal to those who expect arguments concerning the identity of events to run parallel to arguments concerning the identity of continuants. For in the case of continuants, the parallel argument to that we are considering for events is clearly invalid. Suppose that there is a lizard with a tail. Had there been no tail, there would still have been a lizard: indeed it would have been the very same lizard as that which is actual, albeit without a tail. We certainly do not conclude that the actual lizard is a continuant without a tail! Rather we conclude that the tail of the lizard in the actual world is a non-essential part of it. By parity of reasoning we conclude that Monica's having that last drink is a non-essential part of the actual party. And that leaves open the possibility that the rising of the arm is a non-essential part of the actual attempt.

In showing the argument to be invalid we have so far been exploiting the possibility that an event in some possible world may be identical with an actual event though the event in that possible world has different parts from the event in the actual world: a possibility afforded us by the view that events, like continuants, may have non-essential parts. But what if you deny this and hold instead that all parts of events are essential parts? Then the argument is valid.[3] For this view blocks the possibility that a token event in the actual world may occur in some possible world *with different parts.* Given this view, the token attempt which actually occurs would have occurred had the arm not risen only if *all parts* of the actual attempt would have occurred had the arm not risen. Conclusion: the actual attempt does not have arm-rising as a part.[4]

However, even if the argument is valid when offered in conjunction with this view, is it sound? It seems to me that if such a view is correct, then we have good reasons for thinking that the crucial premise, that 'the same [token] event, the attempt, would have occurred had the agent been paralysed', is false. For if this is meant to be knowable *a priori* it amounts to *a priori* physiology. For how are we to know *a priori* that some physical events which occur when an agent tries to raise his arm and raises it, would have occurred had the agent been paralysed? Mightn't paralysis just be a condition such that, when an agent is paralysed, his reasons for acting cause events to occur which they wouldn't have caused to occur had he not been paralysed? (Perhaps *that's* why a paralysed man can't move.) If this is possible, then so much for the crucial premise's *a priori* status.

---

[3] The argument is now only as plausible as is the view that all parts of events are essential parts. According to this view World War Two would not have taken place had some soldier who had a drink during the war instead refused that drink: a *different* war would have occurred. And how plausible is that?

[4] Strictly the argument so construed is of the form 'If x, all of whose parts are essential parts, would have existed had y not existed, then y is no part of x'. Those who would advance such an argument here would deny that we should expect a parallel between arguments concerning the identity of events and continuants on the grounds that, unlike events, not all parts of continuants are essential parts.

Of course, the premise may be offered as an *a posteriori* claim: one whose truth-value we could discover by the investigation of various attempts. But even if it is, I suggest that we should reject it. In order to see why, consider how the argument of the same form given above would run according to this view. Suppose that there is a party and that towards the end of it Monica has a last drink. We may advance the *a posteriori* premise 'Had Monica not had that last drink the same event, *viz.* party, would still have occurred'. Given that all parts of events are essential parts, this premise is false if 'same' is read as 'same token'. What is true, given the essentialist view, is that a *different* party would have occurred had Monica not had that last drink. Thus the reading of the crucial premise which renders it true, where 'same' is read as 'same type', is not one which will allow us to infer that the actual party is an event which does not include Monica's having that last drink. And that seems right. For we know that Monica's having that last drink *is* a part of the actual party. But if this is right, then, likewise, why mightn't the actual attempt be an event which has arm-rising as a part? This cannot be rejected on the grounds that, had the agent been paralysed, the same token attempt would have occurred, and arm-rising would have been no part of *it*. For what is being asked is precisely why we should think that it would have been the same token attempt which would have occurred rather than, like the party which would have occurred, a *different* event. In order to explain this, those who offer the argument under consideration would have to bolster their argument with some theory of attempts according to which arm-rising is no part of the actual attempt. But that would render their argument circular. It would entail its conclusion only if it assumed it to begin with. Moreover, giving a theory of attempts is so closely bound up with giving a theory of action that it is unlikely that those who think that arm-raising has arm-rising as a part, and who accept that all successful attempts are identical with the actions the attempted performance of which was successful, will advance a theory according to which no attempted arm-raising has arm-rising as a part. For they will think that some attempted arm-raisings, those which occur when an agent raises his arm, *do* have arm-rising as a part.

At the outset I assumed that all arguments for the view that tryings are internal events are variations on the argument considered here and shown to be fallacious. If that assumption was a good one, then such arguments provide no obstacle to the view that some attempts are events which include gross bodily movements. And that means that there is no straightforward argument of the kind that Jennifer Hornsby outlines which shows that all actions are internal events.[5]

*Christ Church,*
*Oxford*

© MICHAEL SMITH 1983

---

[5] Thanks to Jennifer Hornsby for comments on a draft of this paper.

# MR MORE ON RIGIDITY AND IDENTITY

*By* MICHAEL LISTON

M. J. MORE (ANALYSIS 42.2, March 1982) argues that Kripke employs two non-equivalent criteria for rigid designation:

(A) 'E' is rigid if 'E might not have been E' is false, i.e. if *being E* is an essential property of E;

(B) 'E' is rigid if it designates the same thing in every possible world in which it designates anything and means E.

While More's argument that (A) and (B) are not equivalent is incontrovertible, it is hardly relevant to Kripke's treatment of rigidity, firstly because Kripke does not hold (B), and secondly – and more importantly – because More seems to misunderstand Kripke's strategic use of intuitive tests to make semantic distinctions.

Although More is not alone in attributing (B) to Kripke,[1] I can find no textual evidence that Kripke holds (B). He uses, rather, a second, closely related, criterion:

(B′) 'E' is rigid if it designates the same thing in every possible world in which that thing exists.[2]

We may, of course, now ask whether a counterexample similar to More's can be constructed which will show the non-equivalence of (A) and (B′). Let us grant that More is right: 'the father of Elizabeth II' is not rigid under (A), since the father of Elizabeth II, George VI, indeed might not have been the father of Elizabeth II, as would have been the case had he lived a life of celibacy. As to whether the description is rigid under (B′), we seem to lack a clear-cut answer. 'The father of Elizabeth II' might be taken to be non-rigid under (B′) in the sense that it does *not* designate George VI in every possible world in which he exists; it fails to designate him (or anything else), for example, in those counterfacual situations in which he remained celibate. Nevertheless, there is also a sense in which 'the father of Elizabeth II' is rigid under (B′) since it designates George VI in every possible world in which he exists whenever it designates anything. (B′), then, may or may not be equivalent to (A), depending upon what convention we adopt as governing the rigidity of definite descriptions in cases where they fail to designate.

---

[1] Hilary Putnam also does, 'Meaning and Reference' in *Naming, Necessity, and Natural Kinds*, ed. Stephen P. Schwartz (Cornell U.P. 1977), p. 128.

[2] 'Identity and Necessity' in Schwartz op. cit., p. 179: 'All I mean is that in any possible world where the object in question does exist ... we use the designator in question to designate that object.'

It seems to me, however, that we should not even be asking whether (A) and (B′) [or (B)] are equivalent, since it is unlikely that Kripke ever intended a rough intuitive test such as (A) to be equivalent to the precisely formulable technical notion of rigidity given by (B′). (A) cannot even be used as a test when 'E' is a definite description, until certain scope distinctions are made! If I understand Kripke correctly, he uses (A) not as a strict test for deciding for any referring expression, 'E', whether it is rigid or non-rigid, but as one of a variety of considerations[3] – modal, epistemological, and semantic – which motivate the thesis that English proper names are rigid *de jure*. In other words, (A) provides a philosophical reason compelling us to accept the claim that the semantic conventions of English are such that proper names behave rigidly, where rigidity is clearly defined in the formal modal semantic terms given by (B′). If this is so, it seems inadvisable to wonder whether (A) and (B′) are equivalent.

*University of California at San Diego*
*La Jolla, California 92093, U.S.A.*

© MICHAEL LISTON 1983

---

[3] See Nathan Salmon, *Reference and Essence* (Princeton U.P. 1981), pp. 23-32 for a discussion of these considerations.

---

# ANOMALOUS MONISM: REPLY TO SMITH

*By* TED HONDERICH

MY paper 'The Argument For Anomalous Monism' (ANALYSIS 42.1, January 1982) advanced the argument that Donald Davidson needs to choose whether it is a mental event as mental or that same event as physical that is causal with respect to an ensuing action, and that either choice is fatal to Anomalous Monism. If it is the mental event as mental that is causal, we have psychophysical lawlike connection, which Anomalous Monism denies. If it is the mental event as physical, we have the inefficacy of the mental, which is intolerable and which Davidson appears to deny.

Peter Smith ('Bad News For Anomalous Monism?' ANALYSIS 42.4, October 1982) claims the same argument was advanced earlier by Frederick Stoutland, and in that he is more or less right. Stoutland's version, of which I was unaware, is in ways inexplicit and involves certain different assumptions with which I disagree, but it is indeed a version of the same argument. My concern here is not to claim sole ownership of the goods but to see whether Smith is right in suggesting they are not worth owning.

He assumes, perhaps too quickly, that Davidson must allow that it is the mental event as physical which is a cause of the action. (It may be, rather, that attempts can be made to resist the argument not by supposing that it is the event as mental which is causal, but by relying in a certain way on the identity-proposition or by denying the propositions that produce the dilemma, notably the proposition about causally-relevant properties or characters.) Having assumed that it is the mental event as physical which is causal, Smith notes correctly that on Davidson's view it is not a matter of lawlike connection that the event as physical *is* the event as mental. That is all that I meant by saying that it is a matter of accident that the physical event is the mental event, and presumably what Stoutland meant by saying the identity is merely a 'brute fact'. The proposition that it is a matter of accident is not, as Smith supposes, an independent proposition got by inference from the proposition denying lawlike connection. To say the same thing again, it is nomologically inessential, to the event's being the physical event it was, that it was the mental event it was. Hence, it is inessential to the relevant effect, the action, that the event was the mental event it was. Here we have epiphenomenalism, the inefficacy of the mental.

Smith's hope of securing the efficacy of the mental rests entirely on one thought. It is that the Anomalous Monist, when asked *which* physical event is to be taken as identical with the mental event, will answer that it is the physical event that has certain causal connections. It stands in causal connection with what ordinary belief ('folk psychology') takes as the explanatory or causal antecedents of the mental event, and the effects of the mental event, notably the relevant action.

Smith's hope is vain, and the answer he supplies to the given question on behalf of the Anomalous Monist does not come near to showing that the argument against Anomalous Monism 'entirely collapses'. The Anomalous Monist, having identified in Smith's way the physical event he has in mind, will presumably persist in the denial of lawlike connection noted above. It therefore remains as inessential as ever to a certain physical event's being as it is, and having the causal connections it does, that it is a mental event to which ordinary belief – but not the Anomalous Monist – assigns the same causal connections. *That* is the objection. It is an objection which does not collapse but rather is quite untouched by the thought as to the identification of the mental event.

To repeat, Smith writes 'it is not an "accident" or "brute fact" that the belief (in virtue of its physical properties) caused the action; for the physical state which is the belief is partially *identified* as the state which has the right physical properties to cause the action!' (op. cit. p. 222). That is a non-sequitur. From the propositions that a state has physical properties which cause an action and which stand in certain other causal relations, and that the state is identical with one commonly believed in virtue of its mental properties to stand in those same relations, and that it is not a matter of lawlike connection that the state as physical is the state as mental, it does not follow that the mental properties of the state are essential to the action.

In the remainder of his paper, Smith considers and attempts to meet other rejoinders to his rescue attempt. I shall not consider this, except to remark that it certainly is mistaken to infer, from the fact that we can establish particular causal claims without knowing laws, that the items might be in causal connection without being in lawlike connection. Indeed, Davidson makes that point (*Essays on Actions and Events*, Oxford 1980, pp. 16–17). Smith's reflections require the inference. Nor shall I consider further and more fundamental difficulties which Anomalous Monism shares with any Identity Theory of the mind (for some of them see my 'Psychophysical Lawlike Connections and Their Problem', *Inquiry*, Vol. 24, No. 3, 1981, and 'Nomological Dualism: Reply to Four Critics', *Inquiry*, Vol. 24, No. 4, 1981).

Given Stoutland's paper, what I announced about Anomalous Monism was not news. If it had been news, it would have been bad news.

*University College,*
*Gower St, London WC1E 6BT*

© TED HONDERICH 1983

---

# MORE ON COUNTERPART THEORY

*By* GRAEME FORBES

IN my 'Canonical Counterpart Theory' (ANALYSIS 42.1, January 1982) I gave a provably correct translation scheme for a standard first-order modal language $L_m$ into the extensional language $L_c$ of first-order counterpart theory. The sense in which the translation scheme is correct is as follows: when the counterpart relation is stipulated to have the same structure as transworld identity, i.e. when it is an equivalence relation such that each possible object has exactly one counterpart at each possible world, every theorem of canonical quantified S5 is translated into a theorem of counterpart theory; and every theorem of counterpart-theory which is reverse-translatable by some $L_m$-sentence is reverse translated by a theorem of canonical S5. 'Canonical' S5 is S5 with actualist quantifiers, contingent existents and strong necessity, and without the Falsehood Convention ('FC'). I call this system 'canonical' since it is the one presented by Kripke in 'Semantical Considerations on Modal Logic' (reprinted in Linsky, ed., *Reference and Modality*, O.U.P. 1971; see footnote 11, p. 66, for a comment on FC), and is the underlying system in Fine's recent work on quantified modal logic, which has the version I use here. The main technical accomplishment of the scheme is to solve the problem of contingent existence, outstanding since Lewis's original paper.

In their 'Reply to Forbes' (ANALYSIS 42.4, October 1982) William Seager and Graeme Hunter raise a philosophical query about my motivation, which I shall answer here, and also put forward three technical objections (one is perhaps just a comment) to my translation scheme, each of which rests on a failure properly to distinguish between strong and weak necessity. So before discussing what they say, I give evaluation clauses (S$\Box$) and (W$\Box$) for these two notions. (S$\Box$) is Fine's clause (iv), 'Model Theory for Modal Logic Part I', *The Journal of Philosophical Logic* 7 (1978), p. 129:

(S$\Box$) $w \models \Box\phi$ iff $v \models \phi$ for all $v \in W$.
(W$\Box$) $w \models \Box\phi$ iff for all $v \in W$, if the referents
of all $t$ in $\phi$ exist at $v$, $v \models \phi$.

For more on weak necessity, see M. K. Davies, 'Weak Necessity and Truth Theories', *The Journal of Philosophical Logic* 7 (1978), pp. 415–39.

## I

Hunter and Seager's main objection is that on my translation scheme, the Falsehood Convention implies that 'nothing can have any essential property without being a necessary existent' (op. cit. p. 226). But this is simply false. The argument they give shows only that nothing can have a necessary property without being a necessary existent. To regiment 'a is essentially F' they choose the formula '$\Box Fa$'; but 'a is essentially F' means, in standard possible worlds discourse, 'a is F in every world in which a exists', and when '$\Box$' is strong, must be formalized by '$\Box(\text{Exists}(a) \rightarrow Fa)$' (thus it is trivial that existence is an essential property, but non-trivial that it is a necessary property). It is true that in his philosophical writings, Kripke uses '$\Box Fa$' to say that a is essentially F, but he is careful to note that in this use he is reading '$\Box$' weakly (see 'Identity and Necessity', in Munitz (ed.) *Identity and Individuation*, New York University Press 1971, p. 137). By my translation scheme, 'a is essentially F' is true just in case for every world w and for every x, if x is a counterpart of a at w then if x exists at w, then x is F at w. The whole point of the scheme is that this condition does not require that a exists at every world.

## II

According to Hunter and Seager, the sentence '$\Box$Rab' is ambiguous. I claimed that the ambiguity arises only because of an expressive weakness in weak necessity, and that with strong '$\Box$', the alleged four readings of '$\Box$Rab' are in fact four different modal sentences. In their 'Reply', they say 'we find it difficult to see how Forbes . . . can produce four $L_m$ sentences with the appropriate semantical qualities' (p. 225). But it is not so difficult:

(1) $\Box$Rab
(2) $\Box$(Exists(a) $\rightarrow$ Rab)
(3) $\Box$(Exists(b) $\rightarrow$ Rab)
(4) $\Box$(Exists(a) & Exists(b) $\rightarrow$ Rab).

## III

In fact, the truth of '$\Box$Fa' requires that a be a necessary existent only if the Falsehood Convention is in force, for otherwise a could satisfy an atomic predicate F even if it did not exist (consider the predicate 'is dead' in the analogous temporal case, where '$\Box$' is ready 'Always'; FC would be inappropriate for tensed languages). Hunter and Seager think that I am forced to impose the Falsehood Convention in order to validate the inference '$\Box$Rab $\vdash \Box(\exists x)$Rax', and they say that there is 'clearly a sense' in which this inference is correct. But I do not agree that the inference is in any way unproblematic. With actualist quantifiers and strong '$\Box$', which many find the natural choices, the claim that atomic predicates must be false of non-existents implies a not wholly pellucid conception of 'simplicity' such that it is analytically true that it could only be existents which have simple properties or stand in simple relations. Note that the inference seems wrong for time: it is always true that I am my father's descendant, but not always true that there exists someone of whom I am a descendant.

What is counterpart theory for? Hunter and Seager say that it is 'to provide an apparatus for expressing modal facts' (p. 225), but insofar as their view that it can express facts or exhibit distinctions not expressible by or exhibitable in ordinary modal language is based on their unclarity about strong necessity, they have yet to provide the theory with a rationale. By the straightforwardly provable correctness result for the translation scheme mentioned at the outset, it is a purely technical fact that counterpart theory, or equivalently, counterpart-theoretic model theory, is required only when we wish to model (or describe) situations in which transworld heirlines are not identity-like (for cases, see my 'Thisness and Vagueness', *Synthese* 1983), unless a rationale for counterpart theory need not manifest itself in terms of which $L_m$ sentences are satisfiable.

The main objection to counterpart theory concerns the necessity of identity, which Hunter and Seager still do not discuss. Kripke has pointed out (op. cit. pp. 135–8) that we can prove

(5) $a = b \rightarrow \Box(a = b)$

using just Leibniz's Law and

(6) $\Box(a = a)$,

the relevant instance of the Law being

(7) $a = b \rightarrow (\Box(a = a) \leftrightarrow \Box(a = b))$.

But in counterpart theory, '$a = b$' does not imply '$\Box(a = b)$', since the truth of the latter requires that at each world w, each counterpart of a at w and each counterpart of b at w are identical, which will fail if at some world, a (i.e. b) has two existing counterparts. Kripke's argument fails in counterpart theory because the translation of (6) is not a theorem if counterparthood is not identity-like. Unfortunately, there is an ambiguity in my specification in 'Canonical Counterpart Theory' of the translation clauses for the modal operators when they have within their scope constants $t_1 \ldots t_n$ not within the scope of some further modal operator itself within the scope of the first. Here the metalinguistic variables range over constant-tokens, of which there are two in (6), even though they are of the same type. Thus the correct translation of (6) is

(8) $(\forall w)(\forall x)(\forall y)(Cxaw \,\&\, Cyaw \rightarrow x = y)$

and not

(9) $(\forall w)(\forall x)(Cxaw \rightarrow x = x)$,

(a scheme which would give (9) for (6) has a number of unacceptable consequences, one of which is that it does not translate (7) into an instance of Leibniz's Law). (8) is a non-theorem since a may have distinct counterparts at w. But this is the price we pay if we think that one object could have been two, or two one, a thought which, in certain special circumstances, I am prepared to defend.

*Tulane University*
*New Orleans, La. 70118, U.S.A.*

© GRAEME FORBES 1983

# IS NATURALISM INESCAPABLE?

*By* JEFFREY GORDON

MANY enemies of theism (Russell, Camus, and Sartre, for example) have shared the predominant view of theists that the meaningfulness of human life requires the existence of God, that a wholly naturalistic universe is a congeries of blind accidents with neither rhyme nor reason, a universe in which man, too, is a pointless eruption in time. In 'The Concept of the Supernatural' (ANALYSIS 37.3, March 1977, pp. 113-6), Gilbert Fulmer argued that precisely this impersonal world is a logical consequence of *theism* no less than it is of atheism. Hence, theism provides no advantage whatsoever in the quest for the intelligibility of the world. If he is right, this should be important news to a great number of thinkers. I want to take a few moments now to reconsider his argument.

Following hints in Hume and Mill, Fulmer argues that theism can provide no alternative to a naturalistic explanation of the world, for the concept of a supernatural creator is incoherent. His argument can be stated simply: Even if nature and her laws were created by God, a 'bodiless rational agent', there is one law of nature he could *not* have created, and this is the law according to which his will is effective: the law that whenever God wills, his will is done. God could not have *willed* this law, since the condition of the effectuality of *any* of his acts of will is that this law obtain. Since God's power to create a world rests upon this law of nature, God's power is itself circumscribed within the natural order. Hence, even a teleological explanation of the universe must revert ultimately to naturalism, and the concept of a deity who transcends nature is hopelessly incoherent. For atheist and theist alike, the world is ultimately impersonal.

Now, the theist might respond that even if he must concede Fulmer's point, there is nevertheless a very great difference between an impersonal world without God and an impersonal world with him. Granting Fulmer's point, one must nevertheless acknowledge an enormous discrepancy between a world guided and sustained by an all-loving Creator and one that is the random product of blind, indifferent forces. When Fulmer writes with the air of one who has vanquished his enemy utterly, he overestimates the power of his point. For it will hardly follow from Fulmer's argument that theism will make no significant difference in one's conception of the universe. Still, even acknowledging the enormous difference between these two conceptions, Fulmer can certainly reply that the difference is not *as great* as the theist takes it to be, that in one crucial particular they are identical: theist and atheist alike must end by embracing naturalism.

So let us go to the heart of the matter: Is this last statement true? Is naturalism the theist's inescapable destiny? I think we will see that it is not.

'Whenever God wills, his will is fulfilled.' Whatever force there is in Fulmer's argument will turn on the aptness of his description of this statement as a law or fact *of nature*. For if it is not precisely a fact of nature, then the theist's teleological account of the world will not be a species of naturalism. So Fulmer's reason for calling it a natural fact will be of crucial importance. And just what is his reason?

> The fact that the god's will is effective is natural in exactly the same sense that any other fact is natural: it occurs independently of any agent's choice. (p. 114)

Since Fulmer specifies no other condition, it is clear that he takes the fact that an event occurs independently of any agent's choice as a sufficient condition of a natural fact. But so far from its being a sufficient condition, the fact that an event occurs independently of any agent's choice is demonstrably *irrelevant* to its being a fact of nature. It is not a *necessary* condition of a natural fact, for if it were it would not be possible for God to will the facts of nature. As the product of an agent's choice, these facts would not be natural. Yet Professor Fulmer seems to have no quarrel with the coherence of the idea of God as creator of the natural world. Would the fact that something occurs independently of any agent's choice be a *sufficient* condition of its being a fact of nature? I think not. Suppose it is the case that every time an angel smiles, the seraphs dance, and that this is not the consequence of any will whatsoever. Shall we feel confident in calling it, therefore, a fact of nature?

The condition named by Fulmer is thus neither necessary nor sufficient to establish that a state of affairs is a fact of nature. Furthermore, adopting his criterion will involve certain serious difficulties. When God supervenes upon the laws of nature, when he turns a woman into salt or commands the waves to part, shall we say these violations of natural law are themselves instances of natural law (namely, the 'natural law' that when God wills, his will is obeyed)? But can natural laws – properly so-called – be thus in contradiction? Surely there must be some way to characterize the invariable effectiveness of God's will that will not commit us to patent absurdities.

Fortunately for the theist there is a ready alternative to Fulmer's odd characterization. The statement 'Whenever God wills, his will is done' is, after all, synonymous with the statement 'God is omnipotent'. If the theist stays with this latter formulation, not even the most zealous naturalist would be tempted to call this a fact of nature, since the statement 'God is omnipotent' is pre-eminently and exclusively a claim about *God*. It would seem to be a minimal condition of a fact of nature that it express a truth about *nature*, but whatever else may be said of a bodiless rational agent, the creator – through an act of will – of rolling sea and starry heaven, it must be agreed on all hands that this is no *natural* being. In other words, there would have to be independent grounds for describing God as a natural being before we could intelligibly describe his omnipotence as a fact of nature. As it is difficult to conceive what grounds these might be, Fulmer's characterization of God's omnipotence as a natural fact must remain unintelligible.

Still, how can we *explain* God's omnipotence? Surely not as the product of God's will, for who other than an omnipotent being would have sufficient power to will omnipotence? Even if my point is granted that God's omnipotence is not a fact of nature, won't the substance of Fulmer's argument rise like Lazarus but in a different form? Won't the theist be forced to end his account of the world not with naturalism but with something just as unwelcome – i.e., the inexplicable fact of God's omnipotence? Why should we prefer this inexplicability to the inexplicability of the laws of nature in the naturalist's account? Even if the theist does not end by embracing naturalism, does he not end, nevertheless, with an ultimately inexplicable universe, precisely what he found objectionable in the naturalist's conception?

But what further explanation of omnipotence is needed than that it is a necessary truth? Of what natural law might this be said? God is in the theist's view a being who possesses every perfection necessarily, a being who cannot *not* be omnipotent. To require an account of God's omnipotence is therefore akin to demanding an account of the number of angles in a triangle. As there is nothing here to be answered, there is nothing legitimate to be asked. The supposed fissure in the theist's world is thus securely sealed.

In the theist's conception of the world, all the facts of nature and all nature's laws are ultimately explained as the product of divine omnipotence. The theist offers this account as an alternative to that of the naturalist, who sees the world as a flash of atoms, with no intention whatsoever in either its origin or process. At the very least, we must recognize a very great difference between these two conceptions. The only question Fulmer's argument forces is whether the teleological explanation must revert in one instance to naturalism, whether the effectiveness of the will of God must itself be construed as a fact of the natural order. Even if we were forced to concede Fulmer's point, the scope of the injury wrought upon theism would not be so great as he supposes. But we are not forced to concede his point. If the teleological conception of the world is indeed incoherent, Fulmer has not shown this. The fulcrum of his argument for the inescapability of naturalism is his characterization of divine omnipotence as a fact of nature. But this characterization is itself incoherent.

*Southwest Texas State University,*
*San Marcos, Texas 78666, U.S.A.*

©JEFFREY GORDON 1983

# RATIONALITY AND PARADOX: A REPLY TO CONEE

*By* FRED KROON

IN 'Utilitarianism and Rationality' (ANALYSIS 42.1, January 1982, pp. 55-9), Earl Conee responds to my 'A Utilitarian Paradox' (ANALYSIS 41.2, March 1981, pp. 105-12) by arguing that a certain difficulty I discerned for utilitarianism can in fact be characterized quite generally as based on a conflict between two apparently compelling dictates of reason:

> $P_1$: While S knows that S's judging x to be W implies that x is R, and knows that being W is incompatible with being R, S is unable to judge that x is W; and similarly with 'R' and 'W' exchanged;

and

> $P_2$: While S knows that S's failing to judge x to be W and failing to judge x to be R implies that x is W, S is unable to refrain from making one of those judgments.

The utilitarian version of this problem has S's judgment of the moral worth (*W*rongness or *R*ightness) of a certain action x contributing to the causal field of that action in such a way that a judgment of wrongness (alt. rightness) would result in the action being right (alt. wrong), and failure to pass judgment would result in its being wrong, with S knowing those various facts.

Conee resolves the conflict between $P_1$ and $P_2$ by arguing that S's rationality never requires the joint satisfaction of $P_1$ and $P_2$. Thus in the kind of problem case under consideration, S may have such a strong rational compunction against knowingly judging falsely that he knows before the interval is over that he will suspend judgment, thus satisfying $P_1$. But this very fact about S also shows that satisfaction of $P_2$ cannot be an *unconditional* demand of rationality. In this kind of case, 'S's rational compunction against knowingly judging falsely wins out over the rational incentive to believe direct implications of what he knows' (Conee, p. 58). Similarly, if S has mere justified belief that he will suspend judgment, and then, in conformity to $P_2$, proceeds right away to draw the MP inference that x must be R (or W), he fails to conform to $P_1$, but does not thereby become irrational: 'drawing such a conclusion would be an instance of a routine sort of fallibility in an extremely peculiar evidential situation ... [and] not a paradoxically irrational lapse' (Conee, p. 58). (These seem to be the interesting cases. I do not discuss the cases discussed on the bottom of p. 58, top of p. 59.)

The solution seems to succeed only at the expense of the idea of a fully rational agent as commonly conceived. A fully rational agent surely does not 'fail to attend to the effect of making an inference', for example. While I grant that conformity to $P_2$ in general is not necessary for rationality – if S dies before he has the chance to draw the inference we should not class him as irrational – conformity to $P_1$ does seem essential to the notion of full rationality. And so, arguably, is conformity to $P_2$ for rational agents under no constraints of time. If conformity to $P_1$ and $P_2$ under these ideal conditions is not seen as *necessary* to the idea of a fully rational agent, then it is difficult to see what is left of the idea. Could a rational agent, on the preferred (weaker) understanding of rationality that tolerates non-conformity to either $P_1$ or $P_2$, also believe an explicit contradiction 'p & ~p' (as opposed to propositions that entail explicit contradictions)? A rational agent could have extremely good reasons for accepting p and extremely good reasons for accepting ~p (in both cases perhaps much better reasons than we normally have for accepting propositions; see, for example, Rescher and Brandom, *The Logic of Inconsistency*, Blackwell, 1980). Should a rational agent, on Conee's understanding of this notion, accept p, accept ~p, and then infer p & ~p? Or should he accept p as well as ~p, and simply refuse to draw the inference to p & ~p? Alternatively should he pay heed to the consequences of accepting p and accepting ~p, and feel rationally bound to suspend judgment concerning p/~p despite the impressive evidence for each of them? Conee would certainly reject the first option (some, of course, accept it, e.g. Richard Routley) and I imagine that he would also reject the second (although some would accept it, e.g. Brandom and Rescher). But if he accepts the third, why not adopt a parallel strategy to deal with the conflict between $P_1$ and $P_2$, namely, have rational agent S suspend judgment about the crucial conditionals that feature in $P_1$ and $P_2$ ('if S were to judge x to be R, then x would be W', etc.) despite the overwhelming evidence in their favour? The reason for not adopting this course of action cannot be that these conditionals are both true, unlike p and ~p, and are perhaps even *known* to be true by others. Rational agents can surely be in possession of reasons for not assenting to some proposition that others can nonetheless have excellent reasons for assenting to, and that others can even be said to know. Nor is it correct to say that what distinguishes the two cases is that suspending judgment about the conditionals featuring in $P_1$ and $P_2$ is positively irrational since there occurs no contrary evidence to weaken the strong evidence each has in its favour. In the case of p/~p too there is no evidence against p or ~p: the contrary evidence concerns only their conjunction.

I do not advocate this strategy, but only mention it to raise some general questions about the idea of rationality as Conee envisages it. The strategy, in fact, is counterintuitive in various ways: thus pursuing it may require us to be sceptical, in the name of rationality, about matters that should be entirely certain. Despite this, I do not recommend Conee's way out. The following example reveals some of the counterintuitive consequences about rationality implied by the strategy above on the one hand and Conee's approach on the other.

Consider sentence C, a self-referential sentence superficially similar to a strengthened version of the Liar but not a genuine semantic paradox (assume that we are working with a non-hierarchical account of truth/falsity such as Kripke's).

> C: Either C is judged false by S when S considers C for the first time, or C is not judged by S to be true or to be false when S considers C for the first time.

Case 1: Suppose that S always conforms to $P_1$ and $P_2$, and consider what happens when he considers C for the first time. On the solution described in the last paragraph, S must rationally accept that at least some apparently undeniable propositions are not in fact true; he must acknowledge that either he is not S, or he is not considering C for the first time, or the conditional 'if S, judging C for the first time, judges it to be false, then C is true' is false (even though this conditional is surely necessarily true), or ..., etc. So the solution of the last paragraph doesn't simply recommend suspending judgment concerning matters that are controversial (e.g. the effects of certain of our actions), but may involve the most evident of matters. Assuming that rational agents should not suspend judgment about such matters, it follows that such an agent S cannot be rational when considering C for the first time. Conee would presumably see this as a reduction to absurdity of the supposition that to be rational an agent must conform to both $P_1$ and $P_2$. I, for the moment, grant only the weaker conclusion that either conformity to $P_1$ and $P_2$ is not necessary to rationality or otherwise rational agents are sometimes led to act irrationally because of special circumstances.

Case 2: Suppose that S is a rational agent who knows that he always conforms to $P_1$, and, in particular, will conform to $P_1$ when considering C for the first time (perhaps he knows this on unimpeachable inductive grounds). Conee certainly grants this as at least a possibility. In that case, however, we have ensured that S knows that he will suspend judgement when considering C for the first time. But we can also assume that S realizes that if he, S, suspends judgment about C when considering C for the first time then the truth of C logically follows. In that case, however, S must fail to draw the logical inference to the truth of C, and hence S does not know C. So if Conee is right C constitutes a counterexample to the thesis that knowledge in the case of rational agents with enough time is closed under known logical implication.

The belief that

$$(1)\quad Kap \wedge Ka(p \rightarrow q)) \rightarrow Kaq$$

is not a valid thesis of an epistemic logic for perfectly rational agents is not a new belief (see F. Dretske, 'Epistemic Operators', *Journal of Philosophy*, 67 (1970), pp. 1007–23; R. Nozick, *Philo-*

*sophical Explanations* (Cambridge, Mass.: Harvard University Press, 1981), ch. 3). But if C is indeed a counterexample to thesis (1), it has one feature not shared by Dretske's or Nozick's counterexamples: it invalidates (1) insofar as it already refutes a much weaker thesis, (2):

(2) $(Kap \wedge Ka(p \rightarrow q)) \rightarrow Baq$

(where 'Ba' means 'a believes').

Thesis (2), unlike thesis (1), is not one that Dretske and Nozick would much care to deny. Now if Conee wishes to deny that (2) is a valid claim about rational knowers (assuming no restrictions of time), it is difficult to see what kind of a priori constraints *do* apply to rational knowers, and how these would be defended. Would Conee similarly reject even a thesis like:

(3) $Ka(p \wedge q) \rightarrow Bap$,

for example? (He should if he agrees that knowledge is at least closed under known logical equivalence, as Nozick thinks, for then it is easy to prove that the rejection of (2) entails the rejection of (3), using the reasoning of Nozick, p. 692, fn. 63. But perhaps Conee would also dismiss the view that knowledge for rational agents is always closed under known logical equivalence.)

It may be thought that example C should not be seen as a counterexample to (1) or (2) for Conee, since S can later – when considering C anew, and hence distanced from his first attempt at judging C – acquire knowledge of the truth of C. Perhaps he can. But this knowledge will not be knowledge culled in virtue of S's rationality but rather in virtue of S's fortuitous circumstances: the fact that he *does* consider C anew, that he is given a second chance. Such an accidental fact scarcely rescues (1) and (2) which are put forward as non-contingent claims. In addition, other cases seem possible where agents whom Conee would admit as rational but who do not conform to $P_2$ can know some conjunction '$p \wedge (p \rightarrow q)$ at any time after $t_0$, but at no time after $t_0$ can instantiate the formula '$Ka(p \wedge (p \rightarrow q)) \wedge Baq$' (consider C′: C′ is either judged false by S whenever S judges C′ after $t_0$, or C′ is not ever judged true/false when S judges C′ after $t_0$, or S sometimes judges C′ true, sometimes false, when judging C′ after $t_0$).

I have tried to draw out some of the implications of Conee's way out, but agree that no rigorous conclusion seems possible. Still, the implications strike me as counterintuitive, and they reinforce my suspicions of Conee's way out and the general theoretical position on which it rests: a 'low' conception of rationality according to which rationality does not make various absolute demands, but is instead a matter of compunctions and incentives that can be safely overriden by other compunctions and incentives. (but what, then, explains their hold on us in cases where they are *not* overridden?) I incline, instead, to an absolutist 'high' conception of rationality, and incline also to the view that in certain situations we cannot do what we rationally ought to do. It is not the rational acceptability of utilitarianism which stands threatened by this possibility, I now acknowledge, but only the supposition that for utilitarianism to be rationally acceptable fully informed and rational utilitarians must always be able to judge in accordance with the demands of their theory.

*University of Auckland*
*Auckland, New Zealand*

© FREDERICK W. KROON 1983

**MIND**

A QUARTERLY REVIEW
OF PHILOSOPHY
edited by D. W. Hamlyn
with the assistance of S. D. Guttenplan

Vol XCII No 367 July 1983

**Articles**
The Disunity of Science: J. DUPRÉ
Race as Merit: M. DAVIS
Trying: O.R. JONES
Conjunctions: Meaning, Truth and Tone: B. RUNDLE

**Discussions**
To Save Fallibilism: J.N. HATTIANGADI
A Cognitive Thoroughfare: M. WELBOURNE
'Egalitarianism': N.M.L. NATHAN
On Some Aspects of Platonic Division: S. MINARDI
Against Epistemic Indolence: F.F. SCHMITT
Self-referent Inference and the Liar Paradox: G.B. KEENE

**Books Reviews**

**Books Received**

*Orders to:*
Journals Department,
Basil Blackwell Publisher Ltd.,
108 Cowley Road,
Oxford OX4 1JF, England

*Annual subscription:*
*Institutions:* £14.50 or $32.50 (Canada $37.75)
*Individuals:* £10.00 or $22.50 (Canada $26.25)
*Single issues:* £4.00 or $9.00 (Canada $10.50)

Published for the MIND ASSOCIATION by BASIL BLACKWELL PUBLISHER

---

*The Philosophical Quarterly*

Basil Blackwell Publisher is proud to announce publication, from January 1983, of *The Philosophical Quarterly*.

MEDICAL ETHICS SPECIAL ISSUE — *Volume 33 no. 132 July 1983*

*In Vitro* Fertilization: The Ethical Issues I — *John Harris*
*In Vitro* Fertilization: The Ethical Issues II — *Mary Warnock*
Gift Relations, Sexual Relations and Freedom — *Loren E. Lomasky*
Foetuses, Famous Violinists and the Right to Continued Aid — *Michael Davis*
The Right to Health Care and the State — *Gary E. Jones*

CRITICAL STUDIES
Bernard Williams, *Moral Luck* — *Raimond Gaita*
Alistair McIntyre, *After Virtue* — *John Casey*

REVIEWS
Recent Work in Political Philosophy — *D.A. Lloyd-Thomas*
Peter Singer, *The Expanding Circle* — *Anthony Manser*
Charles M. Culver & Bernard Gert, *Philosophy in Medicine* — *John Harris*
M. Pabst Battin, *Ethical Issues in Suicide* — *S.E. Marshall*
Robin Barrow, *Injustice, Inequality and Ethics* — *Brenda Cohen*
L.W. Sumner, *Abortion and Moral Theory* — *Stanley S. Kleinberg*
James L. Muyskens, *Moral Problems in Nursing: a Philosophical Investigation* — *R.S. Downie*
T. Regan & D. Van De Veer (edd.), *And Justice for All* — *Antony Flew*
P.S. Atiyah, *Promises, Morals and Law* — *J.P.W. Cartwright*

**Subscription Rates 1983**
Individuals: £9.00 (UK) £9.00 (overseas) $22.00 (USA) $27.00 (Canada)
Institutions: £20.00 (UK) £20.00 (overseas) $45.00 (USA) $55.00 (Canada)
Orders, with remittance, or enquiries to Sue Dommett, at the address below.

**Basil Blackwell Publisher,** 108 Cowley Road, Oxford OX4 1JF

# THEORIA

**A Swedish journal of philosophy**
**Edited by Peter Gärdenfors, Bengt Hansson and Krister Segerberg**

*Volume 48, part 1*
SPECIAL ISSUE ON MORAL PHILOSOPHY
**Sören Halldén** Ingemar Hedenius in memoriam
**Erik Ryding** Are there "archetypical performatives"?
**Carlos E. Alchourrón** and **David Makinson** On the logic of theory change: Contraction functions and their associated revision functions
**John N. Williams** The absurdities of Moore's paradoxes

THEORIA, which invites contributions in all fields of analytic philosophy, usually publishes three issues a year. Subscriptions and requests for back numbers should be sent to *Editor,* THEORIA, *Filosofiska institutionen, Kungshuset i Lundagård, S-223 50 Lund, Sweden.* Subscription for one year is SwCr 60:– or U.S. $ 16.–, for three years SwCr 160:– or U.S. $ 40.–.

# INQUIRY

*Editor: Alastair Hannay*

*Vol. 26 (1983), No. 1*

| | |
|---|---|
| Incommensurability | Harold I. Brown |
| Depth of Intention | Ingemund Gullvåg |
| Scientific Objectivity and the Logics of Science | H. E. Longino |
| Social Science and Social Practice | Francis Schrag |
| It Ain't the Meat, It's the Motion | Richard Sharvy |
| Mathematical Modelling of Election Predictions: Comments to Simon's Reply | Karl Egil Aubert |
| History *v.* Philosophy: the Case of Galileo (M. A. Finocchiaro, *Galileo and the Art of Reasoning: Rhetorical Foundations of Logic and Scientific Method)* | Per Strømholm |

Subscription price:
Institutions: N.kr. 185,— $32.00 — Individuals N.kr. 110,— $19.00.

Published quarterly by Universitetsforlaget, P.O. Box 2959 Tøyen, Oslo 6, Norway — P.O. Box 258, Irvington-on-Hudson, New York 10533, USA.

# NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, J. H. Benson; T. R. Baldwin, Margaret A. Boden, A. E. Fisher, Andrew Harrison, R. F. Holland, Hidé Ishiguro, J. H. McDowell, Bernard Mayo, D. H. Mellor, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

PUBLISHER. ANALYSIS is published by Basil Blackwell Publisher Ltd, 108 Cowley Road, Oxford OX4 1JF, England. US Mailing Agent, Expediters of the Printed Word Ltd, 527 Madison Avenue, Suite 1217, New York, NY 10022, USA. Second Class postage paid at New York.

SUBSCRIPTIONS. The subscription to ANALYSIS for institutions is £11.00 (inland), £13.50 (overseas), $33.00 (USA), $39.50 (Canada); for individuals £7.00 (inland), £8.60 (overseas), $21.00 (USA), $25.00 (Canada). Each volume comprises four numbers, each of at least 48 pages, appearing within the academic year—in January, March, June and October. Orders should be sent to the publisher.

CONTRIBUTIONS. Articles submitted for publication should be addressed to Mr Christopher Kirwan, Exeter College, Oxford OX1 3DP.

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about free copies. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by Whitstable Litho Ltd, Whitstable.

105
Nr 13

ol. 43 No. 4 (New Series No. 200) October 19

# ANALYSIS

Edited by
CHRISTOPHER KIRWAN

## CONTENTS

| | |
|---|---|
| A note on Grim's sorites argument | W. R. ABBOTT |
| Is this a swizzle stick which I see before me? | PATRICK GRIM |
| Ayer's definition of empirical significance revisited | M. L. POKRIEFKA |
| Berry's paradox | DOROTHY GROVER |
| Truth conditions and truth values | ALAN WEIR |
| Nozick on knowledge | B. J. GARRETT |
| Post-hypnotic suggestion and the existence of unconscious mental activity | DONALD LEVY |
| Pettit on revising our understanding of individuals | J. E. TILES |
| The conclusion of practical reasoning | GERALD W. BARNES |
| Is it good to make others do something good? | K. DOŠEN |
| Nagel, Williams and moral luck | JUDITH ANDRE |
| The toxin, the blood donor and the bomb | JANICE THOMAS |
| Human reproduction and sociobiology | JOHN DUPRÉ |
| The coherence of the double standard | JOHN C. BIGELOW *and* ELIZABETH W. PRIOR |
| What was Hume's worry about personal identity? | GEOFFREY SCARRE |
| Conjectures and modalities | BERNARD MAYO |

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003–2638

105

An-13

# SOCIAL PHILOSOPHY & POL

**NEW**

Analysis.

...or and Jeffrey Paul

Vol.

Volume 1, no. 1 September 1983

## Distributive Justice

| | |
|---|---|
| GILBERT HARMAN | Justice and Moral Bargaining |
| CHARLES FRIED | Distributive Justice<br>Comment by **Bruce Ackerman** |
| JAN NARVESON | On Dworkinian Equality<br>Comment by **Ronald Dworkin** |
| WALLACE MATSON | Justice: A Funeral Oration |
| BARUCH A. BRODY | Redistribution Without Egalitarianism<br>Comment by **David Friedman** |
| ERIC MACK | Distributive Justice and the Tensions of Lockeanism |
| DAVID SIDORSKY | Contextualism, Pluralism, and Distributive Justice |
| ANTONY FLEW | Justice: Real or Social? |

**Subscription Rates**

**Volume 1: Autumn 1983 and Spring 1984**

Individuals £10.50 (UK) £12.00 (overseas inc. Eire)
$14.50 (USA) $17.75 (Canada)
Institutions £24.50 (UK) £29.50 (overseas inc. Eire)
$40.00 (USA) $49.00 (Canada)

Orders, with payment, or sample copy requests to: Sue Dommett,
**Basil Blackwell Publisher, 108 Cowley Road, Oxford, OX4 1JF.**

**Basil Blackwell** for the
**Social Philosophy & Policy Center**

ANALYSIS 43.4 OCTOBER 1983

# A NOTE ON GRIM'S SORITES ARGUMENT

By W. R. ABBOTT

PATRICK GRIM (in [3]) has recently argued that sorites arguments are hard to escape. I hope to show that his argument about the wide range of sorites arguments does not succeed.

Grim follows Peter Unger's discussion of swizzle sticks as exemplary of things subject to sorites arguments. Swizzle sticks are existing objects which consist of at least two atoms, and are subject to principle (3):

> (3) If anything is a swizzle stick, then the net removal from it of one atom, or only a few, in a manner most innocuous and favourable, will not mean the difference as to whether or not there is a swizzle stick there. (Unger in [5], p. 519 quoted by Grim in [3], p. 39.)

By repeated applications of (3) it follows that we can get a swizzle stick with only one atom. Therefore the concept of a swizzle stick is incoherent, and there are no swizzle sticks.

Nor will replacing 'swizzle stick' with precise terms like '1-billion-atom swizzle stick' succeed, Grim argues, because anything which is a 1-billion-atom swizzle stick is defined as being both a swizzle stick and as being composed of 1 billion atoms. Since swizzle sticks do not exist, then neither do 1-billion atom swizzle sticks.

But what if we produced a new definition of a type of object by carefully specifying shape, constitution, and number of atoms? Would that avoid principle (3)? Grim argues that it does not, even though the logical inference from something's being a *gizzle gick* (Grim's name for this object) to its being a swizzle stick is not valid. Grim, however, denies that this will avoid the difficulty, because

> (10) If anything is a gizzle gick, that is sufficient for it to be a swizzle stick, in the ordinary sense of the term. (Ibid., p. 42.)

Then (3) above can be rewritten so as to cover 'swizzle sticks, in the ordinary sense of the term', and thus gizzle gicks. In defence of (10), Grim says,

> So were I now to set a gizzle gick, suitably defined, before you, and were I to ask you whether that object would qualify as a swizzle stick in the ordinary sense of the term, I think that you would have to answer 'yes'. (Ibid., p. 42.)

This justification, he thinks, is as strong as the original justifications for (3), which rest on 'linguistic intuitions'.

Grim concludes that a similar difficulty attends any attempt to replace vague terms by more precise ones. In order for precise terms to escape sorites arguments they must apply to the world in such a way that there is no vague term 'F', which also applies to the parts of the world picked out by the precise term. Otherwise, such terms meet the fate of 'gizzle gick'.

Grim's conclusion can be strengthened. There are vague terms such that if any precise terms apply to a part of the world, then at least one of those vague terms does so as well, and thus no precise term escapes sorites arguments. The terms 'object', 'thing', 'process', 'event', and 'state' are vague. Whatever is picked out by a precise term, will be an object, thing, process, event, or state. If sorites arguments are sound, and if Grim's argument is correct, then no terms will escape sorites arguments. Nothing exists. Nor does language exist, or thought, so nothing has been said or thought, if both Unger and Grim are right. Is Grim's argument for (10) sound?

Grim's argument for (10) fails, I believe, because in identifying something as a swizzle stick we are not identifying the same object that is described as a gizzle gick. A simple parallel may clarify the point. When Quine distinguishes (in [1]) between rivers and river stages, he claims that pointing to the region occupied by both the river and the river stage at one time does not fix the reference of the pointing without some rule for fixing the identity of what is referred to. And if you have picked out the river stage, you have not thereby picked out the river.

The moral is that 'imputation of identity is essential, here, to fixing the reference of ostension' (Quine, [1], p. 66). Since gizzle gicks have different identity conditions from swizzle sticks (in respect of subtracting atoms, at least), the two sorts of things stand in a relation to one another roughly analogous to that of river to river-stage. A swizzle stick (if one existed) could overlap a gizzle gick, and then, by a change of atoms, no longer overlap a gizzle gick. Ostension alone does not fix reference. If those who call the object they point to the same object even after some atoms have clearly been subtracted from it, then they are not referring to a gizzle gick but a swizzle stick. So, I conclude, Grim's argument does not succeed.

Not everyone thinks that a difference in identity conditions entails a difference in objects.[1] Suppose, then, that two different terms pick out regions that exactly coincide at time *t*. The two objects will be said to overlap at *t*. If a gizzle gick overlaps a swizzle stick at *t*, then on this theory the gizzle gick is said to be identical with the swizzle stick at *t*. If these objects are identical then they share all predicates. It is, however, clear that gizzle gicks are not subject to principle (3) above, because the removal of even one atom would violate the definition of 'gizzle gick', so that substituting 'gizzle gick', for 'swizzle stick' in (3) will produce a falsehood. Nor do the gizzle gick and the swizzle stick have the same origin, the same future, or the same dispositional properties. What existed in a place qualified as a swizzle stick in the ordinary sense of the term before it qualified as a gizzle gick, and it will continue to qualify as a swizzle stick after the addition or subtraction of an atom disqualifies the application of 'gizzle gick' to that region. The disposition to remain the same object when atoms are either added or subtracted is a property of swizzle sticks but not of gizzle gicks. We cannot avoid this argument by restricting the predicates to be shared by objects identical at a time to those with no reference to past, future or contra-factual conditions. That would show only that identity at *t* was not identity at all. Since no gizzle gick will share *all* predicates with any swizzle stick, it follows that no gizzle gick is identical with any swizzle stick. As for 1-billion-atom swizzle sticks, Grim's argument shows that there are no such things since they are both subject and not subject to principle (3). Since they are swizzle sticks they are subject, and since they contain exactly one billion atoms, they are not. The definition of gizzle gicks is consistent, but that of 1-billion-atom swizzle sticks is not. Therefore, no gizzle gick is a 1-billion-atom swizzle stick. Consequently, the device of gizzle gicks permits precise replacements of our vague terms.

---

[1] The Editor brought this point to my attention.

Some people might reject the above argument. They could respond that the gizzle gick would be the same (i.e. the same swizzle stick) as the swizzle stick at moment *t*. On this view one and the same object has different predicates *qua* different descriptions, and not *qua* others, and thus can be identical without sharing all predicates. But what holds of predicates holds with equal justification of existence. Thus interpreted, Grim's argument shows that the gizzle gick does not exist *qua* swizzle stick, leaving untouched whether it exists *qua* gizzle gick. Consider the story that an early owner of two Gutenberg Bibles was identified as a witch because he possessed two books identical in so many respects. Possessing such unusual objects was paradigmatic of being a witch, we suppose. Calling this person Thomas, we get the following argument:

> Thomas is identical, *qua* witch, with the witch possessing two Bibles.
> Witches do not exist.
> Therefore, Thomas does not exist.

The proper conclusion is that Thomas is not a witch. This may be expressed (awkwardly) as 'Thomas does not exist *qua* witch', if one holds that sort of theory, but the conclusion given clearly does not follow from the premisses. If one objects that Thomas is not identical *qua* witch with anything, since there are no witches, then he should also object that the gizzle gick is not identical *qua* swizzle stick with any swizzle stick, because there are no swizzle sticks. Consequently we see that if the swizzle stick and the gizzle gick are not identical, as Quine argues, then they are identical only *qua* swizzle stick. In either event, we have no good reason for claiming that gizzle gicks do not exist.

*University of Waterloo,*
*Waterloo, Ontario,*
*Canada N2L 3G1*

© W. R. Abbott 1983

## References

[1] W. V. O. Quine, 'Identity, Ostension, and Hypostasis', in *From a Logical Point of View*, New York: Harper, 1963.
[2] W. V. O. Quine, 'What Price Bivalence', *The Journal of Philosophy*, LXXVIII (2), Feb. 1981.
[3] P. Grim, 'What Won't Escape Sorites Arguments', *Analysis*, 42.1 (January 1982), 38-43.
[4] P. Unger, 'There Are No Ordinary Things', *Synthese*, 41 (1979), 117-54.
[5] P. Unger, 'Scepticism and Nihilism', *Nous*, 14 (1980), 517-45.

---

# IS THIS A SWIZZLE STICK WHICH I SEE BEFORE ME?

*By* Patrick Grim

QUINE AND UNGER take sorites arguments seriously against the existence of objects as normal as swizzle sticks, but both suggest precise replacements on the order of '1-billion-atom swizzle sticks' ([3], [4]). W. R. Abbott [1] agrees with me that at least *these* will not do.

But what of more carefully constructed precise replacements such as *gizzle gicks*? With a few hedges these are defined as configurations of plastic molecules within a 100 molecule range of the closest physical approximation to a perfect 5″ swizzloid, a shape patterned on that of an imagined paradigm swizzle stick.

Might not *gizzle gicks* escape sorites arguments? No — or so I originally argued, We start with the existential supposition

(9) There is at least one gizzle gick,

and note that this is inconsistent with:

(10) If anything is a gizzle gick, that is sufficient for it to be a swizzle stick, in the ordinary sense of the term.

(11) If anything is a swizzle stick, in the ordinary sense of the term, then it consists of more than one atom, but only a finite number.

(12) If anything is a swizzle stick, in the ordinary sense of the term, then the net removal from it of one atom, or only a few, in a manner most innocuous and favourable, will not mean the difference as to whether or not there is a swizzle stick there, in the ordinary sense of the term ([2], p. 42).

(10), I argued, rests on linguistic intuititions at least as compelling as those behind Unger's original premises, represented here as (11) and (12). But if we are to maintain (10) through (12) within the bounds of consistency we must also deny (9); taking sorites arguments seriously, there are no gizzle gicks, either.

It is (10) that Abbott attacks.

Suppose that you were to put a gizzle gick, suitably defined, before me now. It's plastic, remember, and quite nearly swizzloid – the ideal shape of an imagined paradigm swizzle stick. It's also blue, perhaps, or a peculiar shade of fuscia, and perfect for stirring bloody marys.

Is this a swizzle stick which I see before me?

'Yes', I say – or at least so I said originally.

'No', says Abbott. Gizzle gicks and swizzle sticks have different identity conditions. A swizzle stick can shed atoms and remain a swizzle stick, but a gizzle gick that loses one atom too many is no longer a gizzle gick. Gizzle gicks are thus not identical to swizzle sticks, and so (10) is false – that this is a gizzle gick is *not* sufficient for it to be a swizzle stick in the ordinary sense of the term.

I have thee not, and yet I see thee still.

Abbott's argument introduces a range of questions regarding identity and existence, and I think his treatment of these is wrong in a number of respects. I also consider these deep and difficult issues, however, and cannot hope to address them adequately here.

Luckily, I don't have to. For identity is simply a red herring – it is not in fact at issue, and the argument can easily be phrased without it.

Unger avoids similar complications of identity by phrasing his sorites arguments without it – (12) above, for example, maintains only that there will be *a* swizzle stick after the removal of an atom, not that the *same* swizzle stick will survive (see also Unger's discussion of avoiding identity in [4], pp. 118-21). Here we can follow suit.

We do not in fact need (10) in the form that Abbott attacks. We need only (10′):

(10′) If anything is a gizzle gick, that is sufficient for there to be a swizzle stick, in the ordinary sense of the term.

No gizzle gick, then, need *be* a swizzle stick. All we need is this: that were I to put a gizzle gick before you now, there would be something before you that would qualify as a swizzle stick in the ordinary sense of the term. With an eye to the definition of gizzle gick, and an appreciation for ordinary notions of swizzle sticks, I don't think that can plausibly be denied.

(10′) still, of course, rests on linguistic intuitions. But these intuitions seem at least as firm as those behind (11) and (12), borrowed from Unger's original. Issues of identity of the sort Abbott raises don't touch (10′) at all.

The argument above with (10′) substituted for (10) has the same force as before. If we are to maintain (10′) through (12) within the bounds of consistency, we must also deny (9); there are no gizzle gicks, either.

If we are to take sorites arguments seriously, then, even precise replacements of *this* sort will fall victim. And identity considerations such as Abbott's will not be enough to save them.

*State University of New York,*
*Stony Brook, N.Y. 11794, U.S.A.*

© PATRICK GRIM 1983

## REFERENCES

[1] W. R. Abbott, 'A Note on Grim's Sorites Argument', *Analysis*, 43.4 (October 1983), 163–5.
[2] P. Grim, 'What Won't Escape Sorites Arguments', *Analysis*, 42.1 (January 1982), 38–43.
[3] W. V. Quine, 'What Price Bivalence?', *Journal of Philosophy*, 77 (1981), 90–5.
[4] P. Unger, 'There Are No Ordinary Things', *Synthese*, 41 (1979), 117–54.

---

# AYER'S DEFINITION OF EMPIRICAL SIGNIFICANCE REVISITED

*By* M. L. POKRIEFKA

AYER'S definition of empirical significance is open to decisive objections by Hempel, Church, Scheffler and Ullian. The definition can be repaired so as to evade these and similar objections. This may come as a surprise to those who believe, with Scheffler, that 'no further restrictions on additional premises for definitions of the Ayer type can reasonably be expected to rule out all conjunctions with nonsensical components' (7, p. 155). Before offering a definition that does rule out all such conjunctions, it will be helpful to review Ayer's definition, some objections, some revisions of the definition, and some more objections.

*Ayer's Definition.* A sentence S is *directly confirmable* iff S is an observation sentence, or there is an observation sentence O and a set of such sentences $\Delta$ such that $\Delta \cup \{S\} \models O$ but $\Delta \not\models O$. S is *indirectly confirmable* iff there is a directly confirmable sentence S′, and a set of analytic, directly confirmable, or independently established indirectly confirmable sentences $\Delta$ such that $\Delta \cup \{S\} \models S'$ and $\Delta \not\models S'$. S is *empirically significant* iff directly or indirectly confirmable (2, p. 13). The definition of direct confirmation is redundant. For if S is an observation sentence, then $\Lambda \cup \{S\} \models S$ and $\Lambda \not\models S$.

*Hempel's Objection.* Let S be any empirically significant sentence and, in all that follows, let N be any sentence. There is a directly confirmable sentence S′ and a set of suitable premises $\Delta$ such that $\Delta \cup \{S\} \models S'$ and $\Delta \not\models S$ . So, $\Delta \cup \{(S\&N)\} \models S'$. Thus, (S&N) is empirically significant (3, pp. 115–16).

*Church's Objection.* Let O1, O2 and O3 be logically independent observation sentences. The sentence: (T) $((\sim O1 \,\&\, O2) \vee (O3 \,\&\, \sim N))$ is directly confirmable with O1 as premise. (T) implies O2 or not. If so, $\sim N$ is directly confirmable with O3 as premise. If not, N is indirectly confirmable with (T) as premise (5, p. 53).

*Nidditch's Revision.* Nidditch's revision eludes the latter horn of Church's dilemma. On this proposal, S is indirectly confirmable iff the set $\Delta$ in the original definition is such that every component of each $P \in \Delta$ is analytic, directly confirmable, or independently established as indirectly confirmable (9, p. 89 which is here construed as in 11, p. 154).

*Scheffler and Ullian's Objections.* Scheffler shows that Nidditch's revision is inadequate. The sence: (Q) $((N \,\&\, O1) \vee \sim O2)$ is directly confirmable with O2 as premise. Hence, N is indirectly confirmable with O1 as premise if $\{O1, O2\} \not\models N$ (11, p. 154). Ullian replaces (Q) in Scheffler's argument with $(N \,\&\, (O1 \vee \sim O2))$ to show that N is indirectly confirmable if just $\{O1\} \not\models N$ (12, p. 275).

Two further objections can be raised against Nidditch's (and Ayer's) definition. First $(N \,\&\, \sim O2)$ is directly confirmable with O2 as premise. Second, $O1 \,\&\, \sim O1)$ is directly confirmable without premises.

*Brown and Watling's Revision.* Brown and Watling sidestep Church's criticism and attempt to improve on a revision by O'Connor. A sentence S is directly or indirectly confirmable on their proposal iff it satisfies a further condition. If S is molecular, then it contains no component whose deletion leaves a statement that (i) entails confirmable sentences not entailed by S or (ii) does not entail confirmable sentences entailed by S. This condition disallows $(O1 \vee O2)$ and $(O1 \,\&\, O2)$ because they contain the sort of components described in (i) and (ii), respectively (4, p. 88). According to Ashby's misstatement of the Brown-Watling proposal, an empirically significant molecular sentence must contain *only* components of the type described in (i) and (ii) (1, pp. 242–43).

Judged on its own merits, this condition fares no better than the condition it misrepresents.

*Cohen's Revision.* Cohen avoids both Church's and Hempel's objections. Let $\phi\&$ be the conjunction of members of the finite set $\phi$ and let 'Cm(S)' be short for 'S is confirmable at level m'. Then: *Cn*(*S*) iff condition 1, 2 or 3 is satisfied.

1. $n = 0$ and S is an observation sentence.
2. $n = 1$ and three further conditions are satisfied.
2.1 S is neither analytic nor contradictory.
2.2 There is an observation sentence O and a set of such sentences $\Delta$ such that $\Delta \cup \{S\} \models O$ and $\Delta \not\models O$.
2.3 For any set $\Gamma$ of sentences: if $\Gamma$ satisfies each of 2.31–2.33, then it also satisfies 2.34.
2.31 $\Gamma\&$ is equivalent to S.
2.32 There is no $\Omega \subset \Gamma$ such that $\Omega\&$ is equivalent to S.
2.33 No extralogical term occurs more frequently in any member of $\Gamma$ than in S.
2.34 There is no $\Gamma' \subset \Gamma$ such that $\Gamma' \cup \Delta \models O$ for each $\Delta$ and O such that $\Delta \cup \{S\} \models O$ and $\Delta \not\models O$.
3. $n > 1$ and three further conditions are satisfied.
3.1 Is the same as 2.1.
3.2–3.3 Are just like 2.2–2.3 except that O can be any sentence confirmable at level $n-1$ and $\Delta$ can be any set of such sentences.
4. *Cn*($\sim$*S*) iff Cn(S).
5. S is *empirically significant* iff Cn(S) for some n (6, pp. 348–49).

Cohen's definition fails to exclude all sentences with nonsensical components. It bestows empirical significance on $((O1 \,\&\, N) \vee (O1 \,\&\, {\sim}N))$ for example (6, p. 349). Moreover, it rules out unobjectionable sentences. Consider the sentences:

(S) $((x)(Fx \rightarrow Gx) \vee ((\exists y)Fy \rightarrow (y)Gy))$

(P) $((x)(Fx \rightarrow Gx) \vee {\sim}((\exists y)Fy \rightarrow (y)Gy))$

(R) $(x)((Fx \,\&\, Fx) \rightarrow (Gx \,\&\, Gx))$

where F and G are observation predicates. (R) is equivalent to the conjunction of the members of {(S), (P)} but none of its proper subsets, no extralogical term occurs more frequently in any member of {(S), (P)} than in (R), and any observation sentence derivable from (R) with suitable premises is derivable from (S) with the same premises (cf. 6, p. 349).

*O'Connor's Revision.* O'Connor's revision is, roughly, that an empirically significant molecular sentence must not contain any component that is not analytic, directly confirmable, or independently established as indirectly confirmable (10, p. 72). This condition rules out sentences with nonsensical components but it does

# ANALYSIS

Edited by Christopher Kirwan

Volume 43

1983

ISSN 0003-2638

BASIL BLACKWELL · OXFORD

# INDEX

| | | PAGE |
|---|---|---|
| Abbott, W. R. | A note on Grim's sorites argument | 161 |
| Andre, Judith | Nagel, Williams and moral luck | 202 |
| Barnes, Gerald W. | The conclusion of practical reasoning | 193 |
| Basu, Dilip K. | Russell on denoting | 65 |
| Bigelow, John C. *and* Prior, Elizabeth W. | The coherence of the double standard | 212 |
| Broughton, Lynne M. | Human reproduction: a self-defeating strategy | 54 |
| Charlton, William | Distance | 19 |
| Cusmariu, Arnold | Translation and belief again | 23 |
| Dale, A. J. | Smith's defence of Lewy | 16 |
| Davies, Martin | Actuality and context dependence II | 128 |
| Došen, K. | Is it good to make others do something good? | 199 |
| Dupré, John | Human reproduction and sociobiology | 210 |
| Forbes, Graeme | Actuality and context dependence I | 123 |
| Forbes, Graeme | More on counterpart theory | 149 |
| Garrett, B. J. | Grayling on internal structure | 78 |
| Garrett, B. J. | Identical truth-conditions | 117 |
| Garrett, B. J. | Nozick on knowledge | 181 |
| Gensler, Harry J. | Acting commits one to ethical beliefs | 40 |
| Gordon, David | Paulsen on the free will defence | 63 |
| Gordon, Jeffrey | Is naturalism inescapable? | 153 |
| Green, Leslie | Dictators and democracies | 58 |
| Grim, Patrick | Is this a swizzle stick which I see before me? | 164 |
| Grover, Dorothy | Berry's paradox | 170 |
| Haldane, John J. | A benign regress | 115 |
| Harrison, Jonathan | Against quantum logic | 83 |
| Hazen, Allen | An example of a language with classical logic for which bivalence cannot be assumed | 1 |
| Helme, Mark | Frege's 'beurteilbarer Inhalt' | 70 |
| Hoag, Robert W. | Mill on conflicting moral obligations | 49 |
| Hollis, Martin | Jim and the Indians | 36 |
| Honderich, Ted | Anomalous monism: reply to Smith | 147 |
| Hornsby, Jennifer | Events that are causings: a response to Lowe | 141 |
| Irvine, Andrew D. | Lucas, Lewis and mechanism — one more time | 94 |
| Jennings, Richard C. | Popper, Tarski and relativism | 118 |

| | | |
|---|---|---|
| Kavka, Gregory S. | The toxin puzzle | 33 |
| Kroon, Fred | Rationality and paradox: a reply to Conee | 156 |
| Levy, Donald | Post-hypnotic suggestion and the existence of unconscious mental activity | 184 |
| Liston, Michael | Mr More on rigidity and identity | 146 |
| Lowe, E. J. | Reply to Hornsby on actions | 140 |
| Maloney, J. Christopher | Dretske on knowledge and information | 25 |
| Martin, Raymond | Tracking Nozick's sceptic: a better method | 28 |
| Mavrodes, George I. | Vestigial modalities | 91 |
| Mayo, Bernard | Conjectures and modalities | 221 |
| Miller, Barry | Cambridge properties and defunct individuals: a reply to C. J. F. Williams | 17 |
| Morris, Thomas V. | Impeccability | 106 |
| Noonan, Harold W. | Personal identity and bodily continuity: a further note on 'The self and the future' | 98 |
| Over, D. E. | Effective and non-effective reference | 85 |
| Persson, Ingmar | Hare on universal prescriptivism and utilitarianism | 43 |
| Pokriefka, M. L. | Ayer's definition of empirical significance revisited | 166 |
| Prior, Elizabeth W. | (see Bigelow, John C.) | |
| Radford, Colin | Report on Analysis problem no. 19 | 113 |
| Rowe, William L. | Self-existence and the cosmological argument | 61 |
| Sainsbury, R. M. | On a Fregean argument for the distinctness of sense and reference | 12 |
| Savigny, Eike v. | Sentence-meaning irreducible to use even within a use-theoretical framework | 5 |
| Scarre, Geoffrey | What was Hume's worry about personal identity? | 217 |
| Sengupta, Gautam | On identifying reference with truth-value | 72 |
| Smith, Michael | Actions, attempts and internal events | 142 |
| Smith, Peter | Smart's argument for realism | 74 |
| Sorensen, Roy A. | Hume's scepticism concerning reports of miracles | 60 |
| Sorensen, Roy A. | Subjective probability and indifference | 15 |

| | | |
|---|---|---|
| Taylor, C. C. W. | A note on ancient attitudes towards slavery | 40 |
| Taylor, Paul | McGinn, token physicalism, and a rejoinder of Woodfield | 80 |
| Thomas, Janice | The toxin, the blood donor and the bomb | 207 |
| Tiles, J. E. | Pettit on revising our understanding of individuals | 189 |
| Waldron, Jeremy | Steiner's vanishing powers | 105 |
| Weir, Alan | Truth conditions and truth values | 176 |
| White, Alan R. | Doubting one's methods | 133 |
| Wright, Crispin | Keeping track of Nozick | 134 |

not rule out sentences with meaningless predicates that do not contain sentence components. Thus, $(x)Hx$ is not ruled out when H is meaningless.

*Another Revision.* Empirical significance must, of course, be defined relative to a language. The present definition is relative to first order extensional languages constructed in the usual way from standard devices. (To simplify matters, it is assumed that the languages contain individual constants.) These languages allow for a rigorous definition and their logical devices are likely to be included in any language for which empirical significance is worth defining. Where $F(a/x)$ denotes the result of replacing all occurrences of free x in F by an occurrence of a:

1. A sentence S is *directly confirmable* iff one of the following conditions is satisfied.
1.1 S is an atomic observation sentence.
1.2 $S = \sim A$ and A is directly confirmable.
1.3 $S = (A \& B)$, $(A \vee B)$, $(A \rightarrow B)$, or $(A \longleftrightarrow B)$ and both A and B are directly confirmable.
1.4 $S = (\exists x)F$ or $(x)F$ and $F(a/x)$ is directly confirmable for some individual constant a.
2. S is *indirectly confirmable* iff one of the following conditions is satisfied.
2.1 S is atomic and there is a directly confirmable sentence $S'$ and a set of directly or independently established indirectly confirmable sentences $\Delta$ such that $\Delta \cup \{S\}$ is logically consistent, $\Delta \cup \{S\} \models S'$ and $\Delta \nvDash S'$.
2.2-2.4 Are like 1.2-1.4 *mutatis mutandis.*
3. S is *empirically significant* iff $\Lambda \nvDash S$ and S is either directly or indirectly confirmable.

None of the objections urged earlier can be raised against this definition. Indeed, a simple induction will verify the following theorem (cf. 8, p. 102).

*Thm.* If S is empirically significant, so are all of its atomic components.

If a sentence is *nonsensical* iff neither logically true, logically false, nor empirically significant, a similar induction shows that:

*Thm.* If S is empirically significant, none of its components is nonsensical.

Whatever other merits or defects they may have, definitions of the Ayer type can escape objections that rely on conjoining an arbitrary sentence with one that already meets the definition.[1]

*Michigan State University,*
*East Lansing, Michigan 48824, U.S.A.*

© M. L. Pokriefka 1983

## REFERENCES

[1] R. W. Ashby, 'Verifiability Principle', in Paul Edwards (ed.), *The Encyclopedia of Philosophy*, Macmillan, New York, 1967.
[2] A. J. Ayer, *Language, Truth, and Logic*, Gollancz, London, 2nd edition, 1946.
[3] ——, *Logical Positivism*, Macmillan, New York, 1959.
[4] R. Brown and J. Watling, 'Amending the Verification Principle', *Analysis* 11.4, March 1951.
[5] A. Church, 'Review of Ayer's *Language, Truth, and Logic*, 2nd Edition', *Journal of Symbolic Logic*, Vol. 14, 1949.
[6] L. Cohen, 'Is a Criterion of Verifiability Possible?', in P. French, T. Uehling Jr. and H. Wettstein (eds.), *Midwest Studies in Philosophy V*, University of Minnesota Press, Minneapolis, 1980.
[7] C. Hempel, 'The Empiricist Criterion of Meaning', *Revue Internationale de Philosophie*, Vol. 4, 1950. Reprinted in 3.
[8] C. Hempel, 'Empiricist Criteria of Cognitive Significance: Problems and Changes', in C. Hempel, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, Macmillan, New York, 1965.
[9] P. Nidditch, 'A Defence of Ayer's Verifiability Principle Against Church's Criticism', *Mind*, Vol. LXX, 1961.
[10] D. J. O'Connor, 'Some Consequences of Professor A. J. Ayer's Verification Principle', *Analysis* 10.3, January 1950.
[11] I. Scheffler, *Anatomy of Inquiry*, Knopf, New York, 1963.
[12] J. Ullian, 'A Note on Scheffler on Nidditch', *Journal of Philosophy*, No. 10, 1965.

---

[1] I would like to thank Richard Hall, Joseph Hanna, Herbert Hendry, and an anonymous referee for helpful criticism.

# BERRY'S PARADOX

*By* DOROTHY GROVER

GIVEN a list of descriptions in English which 'name' integers there will be a least integer which is not named in less than nineteen syllables. It seems that the integer is described by

(1) The least integer not described in less than nineteen syllables.

Suppose we add (1) to the list. Then the candidate referent of (1) fails to satisfy the description. In failing to satisfy the description it can no longer be the referent of (1), but if it's not the referent, it again becomes the candidate referent. This is Berry's paradox. (See Russell 1908. Russell's formulation has 'named' where I have 'described'.)

Berry's description, like the Liar sentence ('This is false'), is usually viewed as posing a challenge to the assumption that crucial characterizing predicates such as 'refers' and 'true' are coherent and consistent. As a result, many philosophers and logicians working with the concepts of truth and reference have sought through the development of formal models to identify (and avoid) contradiction. By contrast, the view I will defend attributes the peculiarities of the Liar and the Berry description to the presence of 'ungrounded inheritors'. According to this account there is no contradiction, for the Berry description fails to refer, and the Liar sentence fails to express a proposition. While this last is not a new suggestion, the first suggestion – that the peculiarities are to be explained in terms of ungrounded inheritors – avoids a recalcitrant problem that current model theorists face. Those who hold both that 'true' and 'false' are characterizing, and that the Liar sentence fails to have a truth value, should also be able to say (consistently) that the Liar sentence is neither true nor false. But they can't. The ungrounded inheritor theorist has an advantage, for that expressibility is neither required nor desired. (This point is made in Grover 1981.) A challenge that remains is the formal identification of ungrounded inheritors, especially those that obstruct communication.

I will begin with a brief summary of the earlier discussions of inheritors. The reader is referred to Grover, Camp and Belnap (1975) (hereafter, GCB) and Grover (1976, 1977) for further details.

## INHERITORS

Successful communication requires a facility for establishing connections between one piece of discourse and another: we need to have some way of communicating that we are continuing with the same topic (when this is the case) or that we are affirming, denying, or considering, something that was said earlier, etc. To establish such connections we use topic and content connecting devices like pronouns. Consider a circumstance where 'she' is used anaphorically in

(2) Tom: Mary will hang the picture later.
Sue: If she does, she'll hang it crookedly.

When context makes it clear that Sue is using 'she' anaphorically we know that Sue intends to refer to the person Tom referred to, Mary. Note that Sue refers to the person Mary, not the word 'Mary'. I have argued elsewhere that natural languages may contain anaphoric devices other than pronouns. There is deletion. Some uses of 'do' seem to qualify as pro*verbs*, and some uses of 'such' as pro*adjectives*. A central claim of the prosentential theory of truth (see, e.g. GCB) is that 'that is true' and 'it is true' function as pro*sentences.*

If we want to make connections with some distant piece of discourse, or with some other discourse, a pronoun (prosentence, or some other kind of proform) will not always suffice, for the anaphoric relation does not ‘stretch’ very far. We can’t use ‘it’ in Chapter 5 of our book to refer to something otherwise mentioned only at the beginning of Chapter 1. There are a variety of expressions that can be used to establish the connections we need. My interest is in those that are similar to pronouns (and other proforms) in that they keep discourse at the ‘object language’ level. I call such expressions *inheritors.*

What if we wish, in this place, Chapter 5, to talk about someone mentioned in Chapter 1? We might say something like

(3) The carrier who delivered Joan’s parcel killed Jim; or
(4) The carrier referred to in Chapter 1 killed Jim.

Donnellan 1966 has drawn our attention to the fact that definite descriptions may function in a variety of ways. My concern is with a possible non-attributive use of descriptions like ‘The carrier referred to in Chapter 1’. (This possibility was suggested to me by some of Kaplan’s work in demonstratives.)

When we want to talk about something mentioned at another time, or in another place, and a pronoun won’t suffice for the job, an alternative available to us is to use a definite description – as in (3) and (4). But if we want to make it clear that we are making connections with some other piece of the discourse, not any kind of definite description will do. Descriptions that contain expressions like ‘referred to’, ‘been talking about’, and ‘just mentioned’, will do the trick, however, for they direct us to linguistic acts. Such definite descriptions are appropriately classified as inheritors for they are similar to anaphoric pronouns in that they establish connections with other parts of a discourse while keeping the discourse at the ‘object language’ level. Consider the use of ‘she’ in (2). ‘She’ refers to Mary, not the word ‘Mary’. And the anaphoric relation connecting ‘she’ and ‘Mary’ is no part of the propositional content of Sue’s response to Tom. Similarly, if ‘The carrier referred to in Chapter 1’ is used as an inheritor in a sentence such as (4), then (4) is about a person (Bill, if the carrier is Bill) and not about pieces of language. The work done by a description in locating an antecedent expression from which a referent is inherited is not part of what is said when a description is used as an inheritor. We can think of that locating work as being done off to the side. In a ‘working column’ perhaps.

Successful use of inheritors requires that they be connected to expressions that acquire their referents ‘independently’ – as a proper name does, for example, perhaps through some causal chain, or as a definite description used attributively does. If a nominal (or sentential) inheritor has as its antecedent something that can in principle supply a referent (or propositional content) it will be said to be *grounded.*

My 1977 paper contained a variety of cases showing how undergroundedness can arise. In the following examples I will italicize the inheritors.

(5) Fay: *The person you are going to refer to* is someone I know.
Burt: No *she* is not.

Assume that both 'The person you are going to refer to' and 'she' are inheritors: each should acquire its referent (if any) from the other; yet each lacks a referent. Circles of inheritors can be quite small, as in

(6) *The person I am now referring to* is coming to dinner.

A language has nominal inheritors *universally available* if, for any given nominal (or nominalizable) expression, an inheritor can be constructed that has the expression as its antecedent. (See Grover 1977). This includes the possibility that an inheritor is its own antecedent, as in (6). Inheritors can be *modified*, as in, 'The carrier not referred to in Chapter 1'. A modified nominal inheritor acquires as its referent (if any) the result of applying some function to the referent of the antecedent.

'Paradoxical' expressions can be constructed when inheritors are both universally available and modifiable. This point can be illustrated in a sketchy review of the ungrounded inheritor account that I have given of the Liar sentence. Suppose $\alpha$ is a sentential inheritor, and suppose we write as a suffix its antecedent. Then in affirming $\alpha_{\text{'Snow is white'}}$ we affirm that snow is white. If inheritors are modifiable by negation, then affirming that $\sim\alpha_{\text{'Snow is white'}}$ will be affirming that snow is not white. We can now envisage a situation in which an inheritor is its own antecedent. This might be represented by $\underset{\curvearrowright}{\alpha}$. The form of the Liar is then $(\sim\underset{\curvearrowright}{\alpha})$.[1] The arrow indicates that $(\sim\alpha)$ is the antecedent of $\alpha$. (This is its form, if we accept the prosentential construal of 'This is false' as a modified prosentence.) On the face of it, an attempt at affirming the Liar would seem to be, not merely affirmation of contradition, but 'affirmation' of incoherence, for we have all-in-one-breath both $\sim\alpha$ and $\sim\sim\alpha$. However, as $\alpha$ is only an inheritor, and ungrounded, $(\sim\underset{\curvearrowright}{\alpha})$ is simply a sentence that fails to express a proposition.

In the next section I will show why I think the Berry description contains an ungrounded (modified) inheritor.

## BERRY'S PARADOX

Suppose we have a list of descriptions that pick out integers. There will be a least integer which is not described in less than nineteen syllables. Suppose we add to the list

---

[1] In my 1977 paper I use $\sim\alpha_{\sim\alpha}$ to give the form of the Liar. This is not an accurate picture because it doesn't show that the subscripted $\alpha$ has $\sim\alpha$ as its antecedent. The arrow is needed.

(1) The least integer not described in less than nineteen syllables.

What is the status of (1)?

We don't have to have something as complex as (1) to generate a 'paradoxical' expression with 'refers'. Suppose we had just a two element domain (1, 2). What is the referent of (7)?

(7) The integer I am not now referring to

On the assumption that the referent is either 1 or 2 it would seem that (7) would have to refer to the other also. My view construes (7), not as embodying contradiction, but as a non-referring ungrounded inheritor: 'not' modifies the inheritor 'The integer I am now referring to'. This suggestion, that the peculiarities of (7) can be accounted for in terms of an ungrounded inheritor, can be made more explicitly another way.

Suppose our language has added to it a sufficient variety of pronouns so that they are universally available. In particular, suppose '*this*' is an anaphoric pronoun that has the largest nominal expression it is contained in as its antecedent. Suppose, also that $\sim$ is a function which yields 2 when applied to 1, and 1 when applied to 2. Then the inheritor reading of (7) can be represented by

(8) $\sim$*this*

This takes us to the familiar $(\sim\underset{\curvearrowleft}{\alpha})$, only this time the inheritor in question is nominal. It is clear that (8) lacks a referent, for the antecedent of '*this*' is (8) itself, and (8) does not acquire a referent independently.

The Berry description is more complex, for what's in question is not just one antecedent, but a set of antecedents. In order to provide an anaphoric pronoun analogue of (1) I will continue to assume that pronouns are universally available. '*That*$_1$, *that*$_2$, . . . , *that*$_n$' will be used as anaphoric pronouns that in the present context have as their antecedents, in order, the less than nineteen syllable descriptions of integers on the original list. An anaphoric pronoun rendering of my inheritor interpretation of (1) is

(9) The least {not{*that*$_1$, *that*$_2$, . . . , *that*$_n$, *that*$_{n+1}$}}

'*That*$_{n+1}$' has (9) as its antecedent. (I have assumed that (9) satisfies the syllable restriction. This could be guaranteed were (9) abbreviated by 'the least {not *that*$_{\{n+1\}}$}'.) 'Not' denotes a function which, applied to a set of integers, yields the complement, and 'the least' denotes a function which, applied to a set of integers, yields the smallest member of the set. '*That*$_{\{n+1\}}$' abbreviates {*that*$_1$, *that*$_2$, . . . , *that*$_{n+1}$}, with '{n+1}' indicating that a set of antecedents is in question.

Like '*that*$_{\{n+1\}}$', 'integer described in less than nineteen syllables' inherits from a set of antecedent expressions. If this fact is not recognized, each of (9) and (1) will be taken to embody contradiction. For suppose we mistakenly assume '$that_{n+1}$' denotes an integer *m*, then (9) will denote something else. But then '$that_{n+1}$' would denote this other integer. Since '$that_{n+1}$' is an inheritor, however, it refers only if it can acquire a referent from (9). But (9) does not acquire a referent independently, since what the referent of (9) is depends on what the referent of '$that_{n+1}$' is. '$That_{n+1}$' is ungrounded, and (9) fails to have a referent. That's the problem with the Berry description: 'integer described in less than 19 syllables' should inherit from a set of antecedents, but as one of the antecedents happens to be the Berry description itself, there's an ungrounded inheritor. The Berry expression fails to refer.[2]

I have used anaphoric pronouns, not just to help articulate the view I am presenting, but to make another point also. '$(\sim\alpha)\!\nearrow$', (8) and (9) show that seemingly paradoxical expressions can be constructed *without* the use of terms like 'true' and 'refers'. This shows that we must look further than models of 'true' and 'refers' in seeking explanations of the peculiarities of the 'semantical paradoxes'. The ungrounded inheritor account is a gesture in that direction, and one that promises a uniform treatment of the semantical and logical paradoxes.

## VARIATIONS ON BERRY

It will occur to the reader that other versions of the Berry paradox may pose more significant challenges for the ungrounded inheritor account. I will examine one case involving the concept of an ungrounded inheritor. My main point will be to show that while there are indeed many things that can go wrong when inheritors are universally available, the cases considered, to date, provide no reason to believe that inconsistency threatens.

Use of a pronoun can be unsuccessful for a variety of reasons. A pronoun, even though it is grounded, may fail to have a referent because its antecedent fails to supply a referent. This happens with 'The king of France is bald, and *he* is wise'. Another problem arises when there is no unique candidate for the antecedent of a pronoun. Yet further difficulties occur when inheritors employ descriptions to locate antecendents. In a case in which a description – rather than an anaphoric relation – locates an antecedent, it can happen that no object satisfies the description. Consider using 'The carrier referred to in Chapter 8' when there is no Chapter 8. In another case, no antecedent is located because the descriptive part of the inheritor contains a contradiction.

---

[2] Is there a problem with my claim that the Berry expression does not refer? If it doesn't refer, an expression such as '$\exists x(x = BE)$', where 'BE' abbreviates the Berry expression, will contain a nondenoting term. I claim there isn't a problem because I reject the inference from "a' doesn't refer' to '$\sim\exists x(x = a)$'. My reasons are given in § 7 of Grover 1981.

In the case of (10) a required set of antecedents fails to be identified.

(10) The least integer not described in a less than 50 syllable description that contains no ungrounded inheritors.

If we continue to assume that an inheritor reading can be represented through the use of anaphoric pronouns, (10) has the form, for some n, of

(11) The least $\{\text{not}\{\mathit{that}_1, \ldots, \mathit{that}_i, \ldots, \mathit{that}_n\}\}$

As with (9), (11) can be abbreviated as 'The least $\{\text{not } \mathit{that}_{\{n\}}\}$'. The anaphoric pronouns of (11) are supposed to have as their antecedents the less than 50 syllable descriptions of integers that contain no ungrounded inheritors.

(11) contains an anaphoric pronoun '$\mathit{that}_i$' just in case the pronoun has as its antecedent a less than 50 syllable description containing no ungrounded inheritors. Will one of the '$\mathit{that}_i$' have (11) as its antecedent? Suppose (11) contains less than 50 syllables. If (11) were the antecedent of one of the pronouns in (11), (11) would contain an ungrounded pronoun. So (11) can't be an antecedent. But if (11) were not an antecedent, all the pronouns in (11) would be grounded, so (11) should be included among the antecedents. Clearly the descriptive condition on antecedents cannot be satisfied. So (11) contains pronouns which don't have antecedents. It fails to refer.

*University of Illinois, Chicago,*
*P.O. Box 4348,*
*Chicago, Illinois 60680, U.S.A.*

© DOROTHY GROVER 1983

# TRUTH CONDITIONS AND TRUTH VALUES

*By* ALAN WEIR

IN a recent paper ('Is this a proof I see before me?', ANALYSIS, 41.3, June 1981, pp. 115–19), Neil Tennant insists that

> We must distinguish a sentence's having a determinate truth value and having a determinate meaning, or truth conditions (p. 116).

and asserts that

> The anti-realist can admit the possibility of definitely meaningful yet undecidable sentences (p. 117).

This is in the course of criticizing Colin McGinn for allegedly misconstruing a Dummettian type anti-realist as denying the above possibility and asserting, in positivistic fashion, that only decidable sentences have a determinate meaning. Tennant's position is that grasp of meaning consists in an ability to decide of a particular purported proof or refutation (in a wide sense of these terms), whether they establish the sentences as true or false. On the other hand, we are not justified in assuming that either the condition for the truth or the condition for the falsity of the sentence obtains, unless we have an effective method for determining which (though it is provably false that neither obtains). Hence determinacy of truth conditions does not imply determinacy of truth value.

McGinn replies to Tennant (ibid. pp. 120–2) that Tennant has completely misunderstood both himself and Dummett. The view that grasp of meaning consists in an ability to effectively decide truth values was not ascribed to Dummett; rather this is Dummett's suggestion as to what a realist would have to say knowledge of meaning, construed truth-conditionally, amounts to. The conclusion that there are no undecidable sentences with a determinate meaning is then taken to be a *reductio ad absurdum* of realist semantics.

It is not clear whether McGinn accepts Tennant's point that anti-realists should allow that sentences with a determinate meaning have determinate truth conditions (at least if they are such as to admit of conclusive verification or falsification). Certainly it would be foolish to reject the point yet ascribe assertibility conditions to the sentence such that all instances of [A* is assertible iff A] held (where A translates the object language sentence A*). For then the realist will be justified in saying that the anti-realist's 'is assertible' is merely a terminological variant of the realist's 'is true'.

In actual fact, however, Dummett sometimes does adopt the position rejected by Tennant, of denying that undecidable sentences have truth conditions (see 'Truth', pp. 17–18, in *Truth and Other Enigmas*, Duckworth, London, 1980). However, it is clear from the context that Dummett is taking it for granted that one who ascribes truth conditions to undecidable sentences is ascribing bivalent truth conditions. In the preface to that book, pp. xxi–xxii, he allows that there is no harm in ascribing truth conditions to undecidable sentences if it is not assumed that it follows that the sentences are either true or false. Thus Tennant's position would seem to be quite consonant with Dummett's views. I wish to make two criticisms of this position.

My first criticism concerns an internal tension within the anti-realist position. For a sentence for which the notion of conclusive verification applies, one can refute a claim that it is neither true (or provable as the anti-realist might alternatively phrase it), nor false (refutable). For a demonstration that the sentence is not true (provable) is a refutation of the sentence: on the anti-realist account a proof that a sentence will never be proved refutes the sentence. Now from this perspective to refute the view that an undecidable sentence has no meaning does not imply that the sentence has a determinate meaning. This only follows if the move from $\sim\sim A$ to A is valid, which it is not, in general, for the anti-realist. The move is valid for decidable sentences, but it is implausible to suppose that an assertion that U expresses a sentence with a determinate meaning in the language of S (that is, from a constructivist perspective, an assertion that there is a particular proposition which one can show to be the one expressed by U), is an assertion of a decidable sentence.

Tennant's position, then, rests on the assumption that grounds for supposing that a speaker can discriminate correct from incorrect purported proofs or refutations of U are grounds for asserting that the speaker attaches a determinate meaning to U. Now it is certainly true that grounds for supposing that the speaker is not disposed to make any intelligible use of U are grounds for denying that the speaker attaches any determinate sense to U. Thus abilities of the sort Tennant alludes to – abilities to decide proofhood of purported proofs – certainly constitute grounds for denying the denial of meaningfulness. But do they constitute grounds for asserting that the sentence has a determinate meaning? What if a more radical anti-realist held that a proof that the speaker could effectively decide the truth value of the sentence provided the only grounds for asserting that the sentence has a determinate meaning. We might describe this position as quasipositivistic. It would not be justifiable to deny meaning to a sentence on the grounds that it was undecidable but, nevertheless, asserting determinacy of meaning would be justifiable just when asserting determinacy of truth value would be: namely when the ability to determine truth value is present. There is a certain internal symmetry about this more radical viewpoint. At any rate we need an argument from Tennant-type anti-realists convincing us that their position is more tenable than the above more radical one.

My second criticism is made from the standpoint of the realist. Suppose we take an undecidable sentence U, such that both realist and anti-realist agree that it is not the case that U is neither true nor false. By a simple application of the double negation rule: $\sim\sim A$ implies A, the realist proves that U is either true or false. What legitimate objection can be made, to this move of the realists? It is certainly not legitimate to object to logicians using fundamental principles of their logic, in attempting to justify that logic. Such a procedure is only viciously circular from the absurd perspective of one trying to derive fundamental principles from independent ones. Nor is justification of one's fundamental principles an impossible or trivial task: for one can attempt to produce a theory of understanding for one's language which certifies that a concept of truth which renders one's favoured logic valid is one which speaker's of the language understand as attaching to the sentences of the language. In doing so it is perfectly licit to use any and all of one's logical principles and rules in the theory of understanding: to deny this would be much more constricting than denying the boxer the use of his hands, or the astronomer his telescope. (Cf. Hilbert 'The Foundations of Mathematics' in van Heijenoort (ed.), *From Frege to Gödel*, Cambridge Mass., 1967, p. 476.) There is no *a priori* reason to suppose that one will be successful, so that the task of producing such a theory is not trivial. (One might describe this task as one of 'naturalized semantics', for it is part of any programme of naturalized epistemology, and an important one at that.)

For the realist, then, given that we can refute the contention that a sentence could never be proved nor refuted, we can assume that one or another of these possibilities is open whether or not we are in a position to effectively determine which one. Hence there can be no objection to ascribing bivalent truth conditions to the sentence. Grasp of these conditions could be explicated exactly as Tennant does in terms of ability to determine proofhood of purported proofs and refutations. (Focus on recognitional abilities seems to me to be a mistake for all but the very simplest sentences, but here I am concerned to show how even accepting anti-realist assumptions about language understanding (with respect to the type of sentence in question), a proponent of classical logic can defend realism against attack.)

Thus it would seem that the anti-realist has no legitimate objection to make to the realist's justification of the classical principle of bivalence. It is common, indeed, for anti-realists to appeal to proof-theoretic considerations, in arguing against realism. Tennant remarks that realists cannot effectively defend instances of $A \vee {\sim}A$, where A is undecidable, and to effectively defend a disjunction is to provide a proof of one of the disjuncts. Others have objected to the failure of classical logic to satisfy the following Conservation Principle which intuitionistic logic satisfies: If A is provable from premisses $\Delta$ then there is a proof of A from $\Delta$ which uses only introduction and elimination rules for the constants occurring in members of $\Delta$ or in A. (A classical counterexample is the provability of $(A \supset B) \vee (A \supset C)$ from $(A \supset (B \vee C)$, which depends on application of a classical negation rule.)

Now such points may bolster an independently arrived at rejection of classical logic, but they cannot provide adequate non question-begging grounds themselves for such rejection. For upholders of bivalence have no need to appeal to effective methods for proving a disjunct, in order to justify the law of excluded middle: appeal to bivalence is as effective a method as one could wish. And, of course, the principle that the truth of the premisses is conserved in the conclusion – the soundness of the inference rules, that is – is the important Conservation Principle for the realist, not the one stated above.

The criticisms above relate solely to sentences for which conclusive verification or falsification is not ruled out. For sentences for which this is not the case – and there are a large number of these – different considerations apply from the point of view of both realist and anti-realist. The notion of assertibility the anti-realist applies to such sentences must be taken to be a defeasible one. Therefore, the anti-realist has to alter the account of negation so that a demonstration that a sentence will never be conclusively proved no longer counts as a refutation. This latter must in these cases require further a (defeasible) warrant for asserting that the sentence will never be warrantedly assertible (in the defeasible sense of the latter notion). A sentence warrantedly assertible at one time may cease to be so at a later time, indeed its negation may become warrantedly assertible. This notion of assertibility could not, then, be taken as a mere terminological variant of the classicist's notion of truth which, for some sentences, is invariant over time. Hence, in this case, the anti-realist's position does amount to a rejection of the idea that truth conditions are applicable to the sentences in question.

The realist, in turn, must, of course, reject the idea that understanding the truth conditions of such absolutely undecidable sentences consists in an ability to recognize those conditions whenever they obtain. But this does not mean that the realist must reject a behaviouristic approach to explanation of speakers' understanding of such sentences. For such sentences will be semantically complex, constructed out of simpler components by means of sentential, and other, operators. If determinate extensions, such as truth functions, can be associated with such operators then, given that the simplest components have determinate truth values, so too will the complex sentences. Understanding of the extensions of the operators will, on a behaviouristic theory, be a matter of appropriately patterning responses to complex sentences with the operator dominant, given responses to the constituents. There is no *a priori* connection between such behaviour and behaviour which would manifest an attempted verification, or falsification, of the sentence. Hence there is no *a priori* reason to suppose that a sentence certified as possessing a determinate truth value be in any way decidable. Whether there are such sentences requires, of course, a much more extended investigation than can be given here.

*University of Edinburgh,*
*David Hume Tower,*
*George Square, Edinburgh EH8 9JX*

© Alan Weir 1983

# NOZICK ON KNOWLEDGE

*By* B. J. GARRETT

IN his recent *Philosophical Explanations* (Oxford University Press, 1981) Robert Nozick has proposed the following insightful analysis of our concept of knowledge:

S knows that p if and only if:

(1) p is true
(2) S believes, via method or way of coming to believe M, that p
(3) If p weren't true and S were to use M to arrive at a belief whether (or not) p, then S wouldn't believe, via M, that p
(4) If p were true (in changed circumstances) and S were to use M to arrive at a belief whether (or not) p, then S would believe, via M, that p.

And Nozick suggests that the possible worlds relevant to the assessment of the subjunctive conditionals in conditions (3) and (4) are those which are (in Lewis's terminology) 'closest to' or 'in the neighbourhood of' the actual world.

However, Raymond Martin ('Tracking Nozick's Sceptic: A Better Method', ANALYSIS, January 1983, pp. 28-33) believes that Nozick's analysis has 'a serious counterintuitive consequence' (p. 31). He gives the following example which purports to show that it is possible for an individual with a true belief that p to satisfy conditions (3) and (4) and yet not know that p:

> Suppose that S bets on a horse at his local track. The bet S places is such that if S picks the winner in the first race or the winner in the second race, or both, S wins ten dollars; otherwise S loses. S picks Gumshoe in the first and Tagalong in the second. But S has the following peculiar method for arriving at his belief whether (or not) Gumshoe wins the first: S shields himself from any information about the results of either of the first two races, and then, after the conclusion of the second race, S presents his ticket to an automated cashier. (The cashier either pays S for his ticker or not, but cannot, due to the way it is constructed, provide S with any additional information about the results of either race.) If the cashier pays S for his ticket; then S believes that Gumshoe won the first; otherwise, S does not believe that Gumshoe won the first. On the occasion under discussion, Gumshoe wins the first and Tagalong finishes last in the second. Tagalong's performance in the second, moreover, was causally independent of Gumshoe's performance in the first: no matter how Gumshoe had performed in the first, Tagalong still would have finished last in the second. After the second race, with no information on the results of either race, S goes to the cashier and presents his ticket. The cashier gives S ten dollars in exchange for S's ticket. The cashier would not give S ten dollars in exchange for S's ticket unless *either* Gumshoe *or* Tagalong had won. But the cashier conveys no information to S about whether it was Gumshoe or Tagalong, or both, that won (pp. 30-31).

Now though S satisfies condition (4) it is not clear that condition (3) is satisfied. Martin holds that it is:

> . . . if it weren't true that Gumshoe won the first and S were to use M to arrive at a belief whether (or not) Gumshoe won the first, then (since the cashier would not have paid S for his ticket) S wouldn't believe, via M, that Gumshoe won the first . . . (p. 31)

But, of course, this is true only if the possible worlds in which Gumshoe does not win the first and Tagalong wins the second are sufficiently 'far' from the actual world as to be irrelevant to the assessment of the claim that S knows that Gumshoe won the first. But if, as Nozick suggests, the interpretation of the tracking conditionals is to be left at the level of intuitive understanding then there appears to be no reason why such worlds should be excluded. Indeed a Nozickian will argue that it is precisely S's insensitivity in such worlds to the fact that Gumshoe does not win the first (i.e. S's failure to track the fact that not-p) which ensures that S's belief does not constitute knowledge. Hence unless we are told why appeal to such worlds in this context is illegitimate we have no reason to believe that Martin has produced a genuine counterexample to Nozick's thesis.

However, consider the following example which is not open to the above objection. Suppose that a person X arrives at a belief about whether (or not) the father of a particular person A is a philosopher on the basis of being told the profession of the father of another person B (in fact A and B are brothers but X is unaware of this). X adopts the following method for arriving at his belief whether (or not) A's father is a philosopher. If B's father is a philosopher (and X is informed of this) then X will believe that A's father is a philosopher. If B's father is not a philosopher then X will not believe that A's father is a philosopher. B's father is indeed a philosopher and X is subsequently informed of this.

In this example all four of Nozick's conditions appear to be satisfied. A's father is a philosopher; X believes that A's father is a philosopher; if A's father were not a philosopher then, since B's father would not have been a philosopher, X would not believe that A's father was a philosopher; if A's father (in changed circumstances) were a philosopher then, since B's father would have been a philosopher, X would believe that A's father was a philosopher. Hence, on Nozick's analysis, X knows that A's father is a philosopher. But this is surely false – it was just good luck that, on this occasion, X's method happened to be applied to two brothers. (Note that I have here assumed the necessity of causal origin – A and B necessarily have the same father.)

However, there are counterexamples to Nozick which do not rely upon any essentialist principles. Suppose that a person Y arrives at a belief about whether (or not) the number of people in Paris is the sum of two odd primes in the following way. If Y is told that the number of people in Paris is even then Y will believe that the number of people in Paris is the sum of two odd primes; if Y is told that the number of people in Paris is not even then Y will not believe that the number of people in Paris is the sum of two odd primes. Further suppose that (necessarily) all and only even numbers are the sum of two odd primes, but that Y does not know this, and that the number of people in Paris is even. Y is informed that the number is even. Here again all four of Nozick's conditions are satisfied. The number of people in Paris is the sum of two odd primes; Y believes that the number of people in Paris is the sum of two odd primes; if the number of people in Paris were not the sum of two odd primes then, since the number of people in Paris would not have been even, Y would not believe that the number of people in Paris was the sum of two odd primes; if the number of people in Paris (in changed circumstances) were the sum of two odd primes then, since the number of people in Paris would have been even, Y would believe that the number of people in Paris was the sum of two odd primes. But, via his method, Y does not know the contingent truth that the number of people in Paris is the sum of two odd primes because he does not know that every large even number is the sum of two odd primes.

Nor can Nozick's analysis satisfactorily account for our knowledge of necessary truths. Nozick holds (p. 186) that conditions (1), (2) and (4) are individually necessary and jointly sufficient for S to know that p, where p is a necessary truth. (Condition (3) is here waived since Nozick is unclear how subjunctive conditionals with a necessarily false antecedent are to be handled.) However, consider the following example. A certain mathematician, Z, believes Fermat's Last Theorem (the as yet undecided conjecture that there is no number n such that $x^n = y^n + z^n$ for positive integers x, y, z and $n > 2$) to be true on the grounds that no counterexample has been discovered. And suppose, for the purposes of this example, that Z's belief is true. Since conditions (1) and (2) are satisfied and condition (3) is not applicable, it follows, on Nozick's analysis, that Z knows that Fermat's Last Theorem is true if condition (4) is satisfied. And condition (4) is indeed satisfied – in any possible world Z would believe, via the same method, that Fermat's Last Theorem is true. But Z does not know that Fermat's Last Theorem is true. His belief to that effect was reached via an unreliable method without the aid of any reliable sub-method.

The consequences of these counterexamples are clear. Unless Nozick can resolve the problems which arise for his theory from certain situations in which a true belief that p is based upon a belief that q (where the believer is unaware of the logical or conceptual relation which holds between p and q) and offer an account of subjunctive conditionals with necessarily false antecedents which excludes beliefs about certain true quantified mathematical statements based upon the no-counterexample method from constituting knowledge, then his unified account of knowledge will have to be supplemented with a condition to the effect that a true belief must be justified or reached via a reliable method if it is known.

*The University,*
*St Andrews, Fife KY16 9AL*

© B. J. Garrett 1983

# POST-HYPNOTIC SUGGESTION AND THE EXISTENCE OF UNCONSCIOUS MENTAL ACTIVITY

*By* Donald Levy

For as long as psychoanalysis has existed, its central concept, that of unconscious mental activity, has been the object of hostile scrutiny by philosophers. Freud replied to such doubts with two sorts of arguments, the first of which he based on the phenomenon of post-hypnotic suggestion. He tended to employ this argument when explaining his views to non-psychoanalysts and beginners in analysis, perhaps just because it seems to establish his fundamental concept without appealing to any of the distinctive clinical data familiar to practitioners of psychoanalysis.[1] One implication of this argument's independence of psychoanalysis is that the psychoanalytic concept of the unconscious will not be undermined if the argument fails. This needs to be emphasized because the main task of this paper is to show that the argument does fail. Here is the argument as Freud states it

> . . . a person is put into a hypnotic state and is subsequently aroused. While he was in the hypnotic state, under the influence of the physician, he was ordered to execute a certain action at a certain fixed moment after his awakening, say half an hour later. He awakes, and seems fully conscious and in his ordinary condition; he has no recollection of his hypnotic state, and yet at the pre-arranged moment there rushes into his mind the impulse to do such and such a thing, and he does it consciously though not knowing why. It seems impossible to give any other description of the phenomenon than to say that the order had been present in the mind of the person in a condition of latency, or had been present unconsciously, until the given moment came, and then had become conscious. But not the whole of it emerged into consciousness: only the conception of the act to be executed – the order, the influence of the physician, the recollection of the hypnotic state, remained unconscious even then.
>
> But we have more to learn from such an experiment. We are led from the purely descriptive to a *dynamic* view of the phenomenon. The idea of

---

[1] See, e.g. 'A Note on the Unconscious in Psycho-Analysis' (1912), *Standard Edition*, volume XII, written for presentation before the Society for Psychical Research of London, and 'Some Elementary Lessons in Psycho-Analysis' (1938), *Standard Edition*, volume XXIII.

> the action ordered in hypnosis not only became an object of consciousness at a certain moment, but the most striking aspect of the fact is that this idea grew *active*; it was translated into action as soon as consciousness became aware of its presence. The real stimulus to the action being the order of the physician, it is hard not to concede that the idea of the physician's order became active, too. Yet this last idea did not reveal itself to consciousness, as did its outcome, the idea of the action; it remained unconscious, and so it was *active and unconscious* at the same time (*Standard Edition*, volume XII, p. 261).

By the end of the first paragraph, the existence of unconscious ideas has been proved only in the trivial sense that forgotten ideas have been shown to exist, which is merely another way of saying that people sometimes forget things they were aware of earlier. Freud's claim at the end of paragraph one

(a) the idea of the order remained unconscious

is equivalent to

(b) the subject did not remember the order.

The crux of the argument really lies in the second paragraph. For if the hypnotic subject did not perform the action ordered, but merely became conscious of the idea of the act, say, then the sense in which the idea of the order was unconscious would be merely that the order had been forgotten, and no more than that. So Freud is correct in emphasizing in paragraph two, that the argument depends upon some idea – the idea of the act ordered – 'becoming active'. Freud sees the need to show that some idea, while remaining forgotten, 'becomes active'. But the idea of the act does not illustrate this, since it becomes conscious, too. So Freud argues

(c) the real stimulus to the act is the order of the physician

therefore,

(d) the idea of the physician's order became active, too.

Unless (d) can be proved, no important sense in which ideas are unconscious has been shown. Does (c) imply (d)? To see why this is doubtful, we need to examine the peculiar meanings of the expressions 'real stimulus' and 'becoming active', when applied to ideas. The former, it will be argued, represents Freud's commitment in this proof to an unacceptable theory of causality, especially mental causality, while the latter expression is ambiguous.

The premise that might at first seem to justify the inference from (c) to (d) becomes clear when we ask what more must be said than

(e) the real stimulus of an act must be active at the same time as the act (or contiguous with it), or must cause something else which is active at the same time as the act (or contiguous with it).

What needs to be added to (e) to justify the inference from (c) to (d) comes out when we realize that the mere time lapse between order and act is not the obstacle to the order being the real stimulus of the act, for Freud. Freud requires that the 'something else' in (e) must be a resemblance of the order, as is the idea of the order. So to capture this view, it seems, we shall need

> (f) the real stimulus of an act must be active at the same time as the act (or contiguous with it), or must cause something resembling the real stimulus, i.e. the idea of the real stimulus, which is active at the same time as the act (or contiguous with it).

(f) would validate the inference from (c) to (d); but is (f) true? It is hard to see why the thing supposed to be caused by the real stimulus must resemble the real stimulus, why it must be an idea of the real stimulus, or why it must be an idea at all. Holding (f) true is like insisting that if the match-flame, that ignites the fuse, that detonates a stick of dynamite, has gone out by the time the explosion occurs, then at the time of the explosion a 'ghost' flame must still really be burning invisibly in its place.

If (f) were accepted, unconscious mental activity could be proved to exist without resorting to so unusual a phenomenon as post-hypnotic suggestion; Freud might more easily have pointed to the simple fact that people obey orders to prove the same thing. We command people to do certain things in the future, and in many cases, at the time indicated, they do as they were told. Must the idea of the order intervene between the giving of the order and the act obeying it? According to (f), it is necessary; people cannot simply do as they were told. However, consider the case in which the subject has long forgotten the order – e.g. a third-grade penmanship teacher may have instructed the subject as a child to cross his t's horizontally, not diagonally. If the adult subject acts in conformity with that order, must we say that each crossing of a 't' now, is necessarily preceded or accompanied by the idea of the teacher's order? (To answer yes is to make the mere occurrence of habit-formation sufficient to prove the existence of unconscious mental activity.) The temptation here is to think that if the idea of the order does not become active, too, then the subject will not know he is to *do* something. He will not know he is to perform the act contained in the idea of the act which also became conscious. The idea of the order, then, explains the meaning of the idea of the act; it indicates to the subject how the idea of the act is to be taken. This is the reason for the demand in (f) that the thing caused by the real stimulus must be the idea of the real stimulus.

But if we need an additional idea, the idea of the order, to tell us what to do with the idea of the act, we shall need yet another idea to tell us what to do with the idea of the order as well. Are we to obey the order, give the order, discuss it, rescind it, disobey it?

The same problem will arise with regard to whatever additional idea is supplied and so on *ad infinitum*. If the idea of the order is supposed to show its meaning unmistakably, so that no further idea is necessary to render its meaning determinate, why is the idea of the act by itself not capable of the same thing, that is, of determining an action with no further idea necessary to make it operative? This infinite regress argument is not intended to show that in cases of post-hypnotic suggestion (or obeying orders, generally) the idea of the order does not become active, either consciously or unconsciously. Its point is only that there is no necessity of the idea of the order becoming active, *a priori*. If Freud's argument were all that we had in support of the claim, we would have no reason at all to believe that the idea of the order becomes active, either consciously or unconsciously, in cases of post-hypnotic suggestion, or any other case.

It might be argued that (f) ought to be treated as an empirical claim of some plausibility, since the best available theories demand that whenever orders are obeyed post-hypnotically, the idea of the order 'becomes active'. Yet Freud offers no empirical evidence of the truth of (f) – his argument is designed to show that the idea of the order *must* become active, if the subject is to be made to perform the act ordered. However, to see why (f) cannot be treated as an empirical hypothesis, we shall have to consider the ambiguity in speaking of an idea 'becoming active'.

When Freud writes of the idea of the act ordered becoming active, he means the subject performs the act. But when he writes of the idea of the order becoming active, this is assumed not to mean that the subject, e.g. *gives* the order; again, the subject obeys the order, since it is the idea of the order which is supposed to activate the idea of the act. Is there any difference in meaning between the two claims – that the idea of the order became active, and the idea of the act became active? Let us suppose that there is no difference in meaning between the two claims; this is plausible since their verification is the same – the subject obeys the order. (f) would then be trivially true; but then, treating (f) as an empirical hypothesis would be absurd. Saying 'the best available theories demand' that (c) entails (d) would be idle.

Suppose, on the other hand, that there is a difference in meaning between the idea of the order becoming active and the idea of the act becoming active. If the difference is of the sort that has already been suggested, i.e. the first gives the sense in which the second is to be taken, then (f) generates an infinite regress, which I take to be a *reductio ad absurdum*. That is, if the reason for supposing a difference in meaning in (f) between the idea of the order becoming active and the idea of the act becoming active is that a guarantee is needed that the occurrence of the idea of the act will result in the *doing* of the act (and I cannot think of another reason for supposing that difference in meaning), then whatever idea is supposed to provide such a guarantee will also stand in need of the very same sort of guarantee; and this infinite regress is a vicious regress. Hence, (f) is either trivial, or leads to a vicious infinite regress.

Assuming (f) to be essential to the post-hypnotic suggestion argument, we can summarize what is objectionable about it this way; there is, in addition to the resemblance requirement it makes, the simultaneity condition, that is, the requirement that one idea be active at the same time as the effect produced by it. What makes this requirement objectionable is first, that the notion of an idea's activity is mysterious, and second, that the notion of the time at which an idea is supposed to be active has been given no sense. What, after all, is the activity of an idea, and when does it take place? Is an idea active merely because we think of it (i.e. if it occurs to us), or only if we also believe the thought it expresses to be true? Must we act on the belief in order for the idea in question to be active? Is the idea in question active if it merely causes some other idea, even if we don't think the second true, and don't act on either? Is the idea active when it occurs to us, or only while we believe it true, or only while we act on it – or just before we do? Can it be active long before, ceasing its activity before its effects begin to appear? These questions are idle, i.e. unanswerable, because there is no way of even trying to determine the truth about them; we have no criterion of when an idea is active, so of course we have no criterion of whether an idea is active at the same time as another idea or event. Insofar as the argument for unconscious mental activity based on post-hypnotic suggestion requires that we treat such questions as meaningful, that argument is unacceptable.

Other arguments, based on recent experimental evidence (e.g. involving selective attention and subliminal perception), support the introduction of the concept of unconscious mental activity without appearing to rely on those aspects of Freud's post-hypnotic suggestion argument criticized here; see 'The Psychological Unconscious: A Necessary Assumption for All Psychological Theory?', by H. Shevrin and S. Dickman, *American Psychologist*, 1980, 35, 5, 421-34. However, a version of (f) seems to be assumed when Shevrin and Dickman define the psychological unconscious as

> that class of psychological events that are at the time unknown to the patient but that actively affect the patient's behaviour (p. 422).

They go on –

> By *active* we mean that unconscious processes affect ongoing behaviour and experience, even though the experiencer may be unaware of this influence. From this point of view, a memory trace indicating some structural modification of the nervous system, would not be considered to belong to the psychological unconscious until such time as it actively influenced psychological events (ibid.).

First of all, the authors are unclear as to what requirement for unconscious mental activity they intend – is active influence of an idea on behaviour necessary, or is such an influence on psychological events alone enough? More importantly, the authors do not state why any 'memory trace' or 'structural modification of the nervous system' which does influence psychological events should be assumed to be active at the time its influence is manifested. Nor do they state why any structural modification of the nervous system, even if it does influence some psychological event, should be assumed to belong to the psychological unconscious at the time its influence is experienced, or be assumed to become a psychological event, or be assumed also to cause an unconscious psychological event, at that time. (Presumably, some chemical processes in the brain, e.g. influence consciousness without the intervention of unconscious mental activities; the mere influence of unconscious brain activity on consciousness ought not be allowed to prove the existence of unconscious mental activity.) Proposition (f) is at work here, it seems, since it would justify these otherwise implausible assumptions.

Let me conclude by emphasizing that nothing I have said implies that unconscious mental activity does not exist; Freud's other proof, based on resistance and transference phenomena, deserves our most careful attention. If there is anything wrong with that other argument, it does not appear to be the sort of thing I have singled out for criticism in the post-hypnotic suggestion argument.[2]

*Brooklyn College,*
*City University of New York,*
*Brooklyn NY 11211, U.S.A.*

© DONALD LEVY 1983

---

[2] I am grateful for criticism and advice received from Richard Boyd, Paul Edwards and Nancy Fish.

---

# PETIT ON REVISING OUR UNDERSTANDING OF INDIVIDUALS

*By* J. E. TILES

THERE are those who have a vision of how science has improved our understanding of the world, which leaves them dissatisfied with our everyday explanations of the behaviour of our fellow human beings. Surveying our practice of identifying the beliefs and desires which give rise to human behaviour, they dream of the day when this 'folk psychology' can be replaced by something 'genuinely scientific'. On the other hand, there are those for whom this dream is a nightmare and who would take comfort in arguments to show our 'folk psychology' could not be replaced or even substantially supplemented. Such comfort can be derived from Davidson's arguments[1] against the possibility of psycho-physical laws, but not by those who feel his arguments rest on an implausible account of how people interpret each others' linguistic acts.

In a recent book, which allies itself generally with Davidson's position, Philip Pettit[2] advances a quite different line of argument, one which draws a protective ring around our practice of explaining behaviour by calling attention to how such explanatory practices differ from those followed in the natural sciences. I wish to argue that 'the remarkable degree of contrast' (p. 99), which Pettit claims to discern, is based on a distorted picture of 'what goes on in our everyday explanation of natural events' (ibid.) I do this not to advance the claims of what Pettit calls 'scientism', but simply to warn those who favour 'humanism' in the human sciences against taking comfort in illusion.

In brief what Pettit aims to defend is the principle (of 'behavioural rationality') that a human being's behaviour issues in a appropriate manner from a pro-attitude toward behaviour with some property and a belief that his behaviour has as it stands a good chance of having this property. This principle, together with the expectation (of 'attitudinal rationality') that the beliefs which accompany pro-attitudes are constrained by, among other things, responsiveness to contradiction and counter-example, constitute the 'orthodox conception of agents' (p. 58).

Pettit in fact draws two lines of defence around this orthodox conception. The inner perimeter (pp. 66–74) guards against the claim that we can *replace* this orthodox conception in a wholesale manner, e.g. by purely neurophysiological explanations of human behaviour. To do this, it is claimed, would leave us unable to interpret the linguistic acts of our fellow human beings and would undermine the difference between the way we treat human beings and the way we treat machines. These arguments seem to me to be on the right lines and I have no wish to examine their strength here.

The outer perimeter is designed to guard against the claim that we can *revise* the orthodox conception, that is, subject it 'to the process of regimentation, experiment and amendment that is characteristic of scientific practice' (p. 74). It is the arguments which consitute this line of defence, which also establish that 'the activity of construing human behaviour in social science cannot be modelled on the parallel attempt to account for natural events' (p. 103). It is the arguments which constitute this line of defence that I wish to examine here.

---

[1] Donald Davidson, *Essays on Actions and Events*, Oxford, Clarendon Press, 1980, Essay 11, 'Mental Events'.

[2] Graham Macdonald and Philip Pettit, *Semantics and Social Science*. London, Routledge & Kegan Paul, 1981. The arguments I shall discuss occur in Chapter 2, 'The understanding of individuals: humanism versus scientism', which is written (see p. 1) by Philip Pettit.

The construction of this line of defence involves a number of claims about scientific explanation which I shall not call into question. The explanation of an event is first of all said to be 'achieved by a statement which advances us in our search for the event's causes' (p. 87). Then it is argued that there is no reason to think this does not apply to people's actions (pp. 89–93). Now, all people offering explanations are committed to there being a regularity which is instantiated by the cause relation which they have identified (p. 94). Explanations are then divided into those where the regularity in question involves the postulation of a law and those where it does not (pp. 93–5). In the former case the explanation is 'nomothetic' and the person offering it is committed to something independently falsifiable.

Not all the explanatory principles of natural science are in this sense nomothetic. The transitivity of the length of rigid bodies and the principle of mass-energy conservation are in different respects not subject to falsification, but have the status of contextually *a priori* principles (p. 96). They are, if not totally then very largely, resistant to being 'rationally denied under the pressure of empirical evidence'; they will be given up only after the adoption of a 'revolutionary perspective' (p. 95). Explanation appealing to such principles is 'non-nomothetic', but as a rule the explanation of natural events is nomothetic.

By contrast, 'the intentional explanation of action is not nomothetic' (p. 93). This means that what is appealed to when an action is explained must be something which is not independently falsifiable. There are no 'circumstances which if they were realized would give us reason to consider revising or rejecting that proposition' (p. 95). So the 'body of principles' applied in 'intentional explanation' is never used in a way which exposes any member to revision, but that body is none other than 'the orthodox conceptions of agents' (p. 101).

This argument provides the basis for what Pettit calls 'the remarkable degree of contrast between what goes on here [in action explanation] and what goes on in everyday explanation of natural events' (p. 99). Making out this 'remarkable degree of contrast' is very important to Pettit's general position. Something like it is needed, since he wants to be able to associate his overall position regarding the understanding of individuals with the *Verstehen* of Dilthey and to claim to be advocating 'humanism' as opposed to 'scientism' (pp. 56, 74)

The crucial thesis is that intentional explanation is non-nomothetic and cannot be nomothetic like explanation in natural science, because intentional explanatory principles are not open to empirical revision. To assess the strength of this thesis we must be precise about what it is that is not to be revised, because at one level, at least, intentional explanations *are* revisable. I attribute a man's gift to charity to a desire to help the needy. When I know him and the pattern of his donations better, I explain what he did as arising from a desire to keep the favour of certain local politicians. But this revision appears to illustrate what Pettit means by 'non-nomothetic': I do not ever find evidence to persuade me to abandon the principle that desire to help the needy may cause a man to donate to charity, I revise *only* by shifting to a different kind of cause. When I alter my assessment of the causal influence, it is never because a law has been refuted, it is only because the wrong cause was identified in the first place.

This looks like a powerful consideration until one examines closely the claim that this is not also characteristic of 'the everyday explanation of natural events'. Admitting that natural science is non-nomothetic at a very high level of generality (mass-energy conservation, transitivity of length of rigid bodies) obscures matters somewhat. In the everyday explanation of natural events the established laws of science are not called into question. When we revise our explanation of a particular natural event, it is not because a law has been refuted or modified, but because we think we have wrongly identified the sort of cause operating, just as I imagined in the previous paragraph wrongly identifying the cause of a man's gift to charity. ('Nothing smashed into it; that is the effect of frost.' 'There is no glue; it is held there by magnetism.')

It is true that behind these explanations we have come to see the workings of natural laws and that similar laws covering intentional actions have not been identified. The issue, however, is whether we have reason to think such laws *cannot* be identified. What is supposed to encourage us to think they cannot be identified is the suggestion that when we resist an explanation of a natural event, we call into question the law upon which it is supposed to be based. But clearly one can reject such an explanation without in any way urging a modification of what we take to be the laws of nature. To use a single case or class of cases to call a law of nature into question requires an elaborate and sophisticated argument concerning how best to explain such cases in the light of our success or lack of success in explaining a vast array of other phenomena.

According to Pettit's own account of explanation (pp. 81–7), a causal relation can be identified without the postulation of a covering law – this in the natural sciences as much as in the human sciences. A very likely account of pre-scientific explanation of natural events, then, is that people identified a variety of causes operating in nature (impact motion, frost, handling toads, the whims of Zeus) without postulating laws, and then chose from among these causes when explaining, and revised only what they thought to be the operating causal influence in some particular instance. The organization of these causal influences into testable theoretical structures based on laws is a late sophistication, and although we now see such laws lurking behind the events we explain, our revision of everyday explanations no more call into question the principles following from these laws, than they do the principle that a desire to help the needy may explain a donation to charity.

It is not, therefore, impossible that as we regimented our explanations of natural events, so we might regiment our intentional explanations. One step in this direction was to consider rejecting the claims of hitherto recognized species of natural cause (handling toads, Olympian deities) to exert genuine influence. On the behavioural side, the claim that no one ever acts from other-directed motives is an intelligible, if false, attempt to do this. Another step was the discovery of hitherto unrecognized classes of causal influences, e.g. radiation, endogenous opiates. Freudian psychology, whatever the extent of its success, was an attempt to do this in behavioural science.

There is a familiar plea made on behalf of the behavioural sciences that they are still in their infancy and may yet grow into something every bit as scientific as modern physics. Tired though this plea may be, Pettit has not laid down an argument to show that in fact the behavioural sciences *can never* (and consequently should not aspire to) grow up in this way. He has merely pointed to a feature of intentional explanation, which at first appears not to be shared by natural explanation, but in fact is shared. Those who advocate 'scientism' in the human sciences might well accuse Pettit of obscuring this shared feature in order to foster a Peter Pan mentality in those disciplines. I would not myself like to be counted among the advocates of scientism. My complaint is only that Pettit has not equipped explanatory dualists like myself with a serviceable distinction between the enterprises of explaining natural and human affairs.

*University of Reading,*
*Whiteknights, Reading RG6 2AA*

©J. E. TILES 1983

---

# THE CONCLUSION OF PRACTICAL REASONING

*By* GERALD W. BARNES

I AM concerned here with the claim that on those occasions when an action is the result of practical reasoning, the action is the *conclusion* of the practical reasoning.[1] Many cases of practical reasoning seem to be of the following sort: *Stage 1:* through a process of considering his reasons for and against bringing about *B*, *S* reaches the belief that it is desirable to bring about *B*, and through a process of theoretical reasoning, *S* reaches the belief that by doing *A* he will, or may, bring about *B*; *Stage 2: S* does *A* for a reason, his reason being his belief that it is desirable to bring about *B* together with his belief that by doing *A* he will, or may, bring about *B*. The claim under consideration is at least difficult to understand otherwise than as requiring that some element of the practical reasoning stand, in such a case, to *S*'s doing *A* as the immediate premise(s) to that alleged conclusion. The most natural suggestion is perhaps that *S*'s reason for doing *A* stands thus to his doing *A*.

---

[1] This claim is said to derive from Aristotle. It has recently been endorsed by, for example, Anscombe, *Intention* (2nd ed.; Ithaca: Cornell University Press, 1969), sec. 33, Davidson, 'How Is Weakness Of The Will Possible?', *Essays on Action and Events* (Oxford: Clarendon Press, 1980), p. 39, and T. Nagel, *The Possibility of Altruism* (Oxford: Clarendon Press, 1970), p. 21. I am indebted to Annette Barnes and Mort Winston for help with versions of this paper.

*S*'s reason for doing *A* does not, of course, stand *exactly* to his doing *A* as premises do to conclusion in standard indicative inference, since doing *A* need not involve a use of language. The view I shall argue for is that there is a crucial defect in any supposed *analogy* here, when *S* does *A* for a reason in a case of the sort described.

At a superficial level, the view I am defending can seem counter-intuitive. One believes something for a reason as well as doing something for a reason; one bases one's beliefs, as well as one's actions, upon one's reasons. Basing beliefs upon reasons *can* often be regarded as a matter of basing a conclusion upon, or inferring a conclusion from, premises. It might be thought that the same or something similar must hold for actions, since the indicated ways of talking run so closely parallel.

Nevertheless, I shall argue that this thought is mistaken. My argument will have the following general structure. Cases where *S* believes for a reason can sometimes exhibit a certain feature. That they can do this is a necessary condition of the possibility of regarding *S*'s believing for a reason as a matter of *S*'s basing a conclusion upon premises. It is also a necessary condition of the possibility of similarly regarding *S*'s *acting* for a reason that cases where *S acts* for a reason be able sometimes to exhibit this or an analogous feature. Such cases cannot exhibit such a feature, however, and so the view I am defending prevails.

Before I elaborate on this outline of an argument, let me introduce a certain restriction upon the discussion in this paper. It seems to me that it is at least possible that the generic characterization, '*S*'s believing for a reason', applies to two importantly different sorts of case. As I see these sorts of case, all and only the cases of one sort can be viewed as cases where *S* infers his belief from his reason. The existence of cases that can be so viewed seems unproblematic, as I have already suggested, and it is exclusively to these cases that my remarks about *S*'s believing for a reason are intended to apply.

With our focus thus restricted, we may begin by considering the following case. *S* is convinced that he will win the raffle; his reason for this belief is a pair of further beliefs: that he has bought up every one of the raffle tickets, and that anyone who does this is virtually sure to win. There are various circumstances in which *S*'s believing for this reason that he will win might be held to be *unreasonable*. It might, perhaps, be held to be so if, although *S* has plausible-seeming grounds for supposing that he has bought all the tickets, the fact is that he has bought only three-quarters of them, the other one-quarter remaining in someone else's possession. Or, we might so judge if *S* has only very poor grounds for supposing himself to have acquired all the tickets. Again, we might so judge if *S* has, and is overlooking, good evidence that he will *not* win the raffle – he knows, say, that the raffle organizer intends to abscond with the proceeds before the drawing.

Such judgments of unreasonableness as those indicated have something in common: the basis of each is some state of affairs that *need not* obtain when *S* believes for the reason in question that he will win the raffle. We can imagine circumstances in which *S*'s believing this for that reason would not be unreasonable at all.

Consider now a second case. *S* is completely convinced that he will win the raffle, on the basis of no other *reason* than *this* pair of beliefs: that he has bought up roughly half the tickets, and that in such a situation one has roughly a 50–50 chance of winning. We are not supposing that *S* believes just that he has a better chance than anyone else, or that he holds tentatively or weakly the belief that he will win; his belief is the firm belief that he *will* win. To make this a realistic case, suppose that beyond *S*'s reason for holding this belief there are additional reasons why he holds it, such as his intense desire to win *this* raffle, his propensity toward excessive optimism, and so on.

Quite apart from any further specification of the details of the case, it would be agreed, I think, that *S*'s believing merely for this reason that he *will* win is in a certain respect unreasonable. Two features of this case are important for our purposes. First, *what* is unreasonable *about* what *S* does here is clearly that in basing the belief that he will win on the reason in question, *S* does something that can be viewed as *inferring* that belief from the reason-beliefs by means of a deductively invalid and inductively weak argument. Further, his basing *that* belief merely on *that* reason cannot fail to exhibit this defect. Hence, second, the basis for the relevant judgment of unreasonableness is *not* any state of affairs that might fail to obtain when *S* believes for this reason that he will win the raffle. There are no circumstances in which *S*'s believing this for that reason would not be unreasonable *at all*, even though *S* might in some circumstances have a reason, even an overriding reason, to be, or to allow or to cause himself to be, thus unreasonable. *S*'s holding this belief for that reason is '*internally unreasonable*'.

Here is a third case. In response to the announcement of a job opening, *S* sends his curriculum vitae to a certain institution, his reason for doing so being this pair of beliefs: that it would be desirable for him to get a job with that institution, and that sending his c.v. there will under the circumstances have a slight chance of leading to that result. There are various ways in which *S*'s doing this for this reason might be unreasonable. One of the beliefs may be, although plausible, false, or, although true, unwarranted for *S*. *S* may have, and be overlooking, weightier reasons for not sending his c.v. to that particular, or to any, institution. There can be, however, no basis for judging *S*'s doing this for this reason to be unreasonable other than some state of affairs that *need not* obtain when *S* submits his c.v. for this reason. *S*'s doing this for this reason is not *internally* unreasonable, for it may, under some at least logically possible circumstances, be not unreasonable at all. One example of such circumstances could be, roughly speaking, circumstances in which *S* knows that he has this and only this chance of advancing whatever purposes would be advanced by his obtaining a job with the institution in question, and in which *S* knows that neither he nor anyone else has in any sense *anything* to lose by *S*'s making the attempt.

What holds for this case can, I think, be seen to hold quite universally for cases of the sort identified at the outset involving *S*'s doing *A* for a reason. There are grounds for supposing that, among such cases, internally unreasonable cases are (with a certain qualification) not *possible*. The qualification is just this: we set aside from the present discussion cases of believing or acting for a reason where the reason includes a belief whose content is (perhaps subtly) self-contradictory or nonsensical. This qualification simplifies the discussion without, I think, undermining its conclusion.

To come to see the impossibility in question, let us note, first, that in cases of the sort we are considering, because part of *S*'s reason for doing *A* is his belief that *by doing A* he will or may bring about *B*, *S*'s reason, *R*, for doing *A* can be a reason *only* for doing *A*, i.e., for performing an act of the description '*A*'. Accordingly, when *S* has reason *R* for doing *A*, *S* can act for a reason otherwise than by doing *A* for *R* only if *S* has some reason, $R'$, different from *R*, for doing either *A* or something else, $A'$ ($A'$ being perhaps only a refraining).

*S*'s having *R* for doing *A* is logically compatible with *S*'s having no such further reason, $R'$; although this sort of situation may not be common, life's opportunities are transitory, and so it is at least a logically possible one. *S* may for the moment have, for example, *no* other reason *for* sending his c.v., and *no* reason for doing *anything* else (e.g., refraining from sending the c.v., or trying some action other than sending the c.v. as a way of obtaining the job).

Further, in the sort of situation just described *S* may be, in a certain sense, 'blameless'. *S* is so, in such a situation, when both in holding the beliefs that comprise *R* and in not holding beliefs constituting any reason $R'$ of the sort specified *S* is being entirely reasonable.

Having reason *R* for doing *A*, lacking any reason *R′* for doing either *A* or any *A′*, and being blameless, *S* may still not do *A* for *R*; *S* may just refrain for no reason from doing *A*, for example. Suppose, however, that in this situation *S* *does* do *A* for *R*. *S* was faced, blamelessly, with these two alternatives: do *A* for *R*, or act for no reason. *S* had, *ex hypothesi*, no reason to act for no reason. However minor the considerations involved in *R*, *S* cannot, I think, be held to have been unreasonable here in doing *A* for *R*.

It seems, therefore, that there are possible circumstances in which *S*'s doing *A* for *R* would altogether avoid being unreasonable. This holds not only when *A* and *R* are as we have taken them to be in the case of *S*'s c.v., but for all possible paired values of *A* and *R*, including those that define cases whose natural circumstances would always provide *S* also with conclusive grounds for not doing *A*, or for rejecting one of the beliefs comprising *R*. Suitable alternative circumstances for such cases will always be at least possible. *S*'s doing *A* for *R* is *internally* unreasonable, however, only if there are *no* possible circumstances in which *S*'s doing *A* for *R* would altogether avoid being unreasonable. Since my argument has not relied upon any special values of *A* and *R*, what seems to follow is that, as I claimed above, it is not possible for *S*'s doing *A* for a reason to be internally unreasonable.

This situation stands in contrast to what we have found in connection with believing for a reason. In the second of the three cases above, *S*'s performance was internally unreasonable. That it could be so is due to (*inter alia*) the fact that any reason of *S*'s for believing that so-and-so is potentially a reason for various *other* beliefs on *S*'s part. Believing for a reason differs here from doing for a reason, since, as we saw, *S*'s reason for doing *A* can be a reason *only* for doing *A*.

There is a certain point of view – exemplified to at least some extent by, for instance, what Descartes says about error in the Fourth Mediation – from which it appears that any *R* that *S* has for believing that so-and-so is capable of being his reason for believing, instead, that it is almost certain that so-and-so, or that it is likely that so-and-so, or that there is a real chance that so-and-so. (This could be reformulated in terms of the notion of degrees of belief.) Which, if any, of these things *S* comes to believe for *R* will depend upon what, if anything, *S* chooses to believe for *R*.

*S*'s doing *A* for *R* is logically compatible with *S*'s having no further reason, *R′*, for doing *A* or any *A′*. Suppose that it could be made out that *S*'s believing that so-and-so for *R* is logically compatible with *S*'s having no *R′* for believing that so-and-so or for believing anything else. Even here, *S* can still believe something for a reason without believing that so-and-so for *R*, if the suggestion of the paragraph immediately above is correct. Nor need *S*'s alternatives here be equally reasonable; there is often a difference of reasonableness between *S*'s believing, for *R*, that so-and-so, and *S*'s believing, for *R*, e.g., that there is a real chance that so-and-so. For many values of '*R*' and 'so-and-so', this inequality of reasonableness remains even on the assumption that *S* is in the sense explained above blameless in connection with his relevant beliefs.

It seems inescapable, then, that *S*'s believing for a reason can be internally unreasonable. It will be so when *S*'s doing so is unreasonable, and the unreasonableness stems from the mere fact that *S* believes *that* for that reason rather than, as it was open to *S* to do, believing a certain something *else* for the very same reason.

I have been trying to explain why I think that internally unreasonable cases of *S*'s doing something for a reason are not possible. Let us now try to see what this thought implies about the conclusion of practical reasoning.

*S*'s believing something for a reason is a type of activity on *S*'s part. It seems possible, I have said, to regard this type of activity as a matter of *S*'s *inferring* what he believes from his reason for believing it, or basing what he believes upon this reason as a *conclusion* is based upon premises.

It is, I think a major element of our notion of *inferring* that a type of activity can be correctly regarded as a matter of inferring a conclusion from premises only if it is possible for *S* to carry out activities of this type in such a way that *S* can be deemed to have inferred what he infers by means of an argument that is both deductively invalid and inductively weak. There are counterexamples to this claim, of course – e.g., if *S* is infallible, or if the type of activity is the inferring of conclusions from premises by means of valid arguments – but if we set aside unrealistic values of *S*, and *ad hoc* types of activity, as I shall suppose we can, then the claim seems relatively uncontroversial. It is simply an application of the idea that what can be done right can be done wrong.

If one type of activity is to count as a good *analogue* of another, then the major conceptual constraints upon the occurrence of activity of the first type should be at least analogous to those governing the second. (Call this the principle of analogous constraints.) Thus, for a given type of activity to count as a good analogue of inferring, it must be possible for *S* to carry out activities of that type in such a way that *S* can be deemed to have done, at a minimum, something analogous to inferring a conclusion by means of an argument that is invalid and weak.

Reflection upon cases like the second of the three examined earlier suggests that internal unreasonableness is an inevitable, conceptually guaranteed, aspect of *S*'s inferring on those occasions when he infers by means of an invalid and weak argument. Inferring by means of such an argument is, by definition, in a certain respect unreasonable. No alteration of *circumstances*, as opposed to an alteration of the *argument*, can make an invalid and weak argument other than invalid and weak. The unreasonableness seems, then, to be internal to such a case, given that cases are being individuated in terms of *S*'s belief and his reason for it.

By the principle of analogous constraints, therefore, for *S*'s acting for a reason to count as a good analogue of inferring, there must be possible cases where, in the first place, in acting for a reason, *S* does something at least analogous to inferring his action from his reason for it by means of an invalid and weak argument, and in the second place, *S*'s acting for that reason exhibits at least an analogue of internal unreasonableness. I have tried to show that there are no possible cases where *S*'s acting for a reason exhibits internal unreasonableness. That some such cases might exhibit an *analogue* of this seems to me to be an entirely unsupportable conjecture.

Unless support for that conjecture can be found, the claim that action is the *conclusion* of practical reasoning is in difficulty, if that claim is meant to imply that some element of the practical reasoning stands as (something like) the immediate premise(s) to the action as (something like) conclusion. The endangered claim fares no better, I suggest, if we replace *S*'s belief that it is desirable to bring about *B* with a desire, or an intention, to bring about *B*. Although my arguments here have not dealt with all conceivable views on the matter, the correct view seems to me to be that there is in practical reasoning no reasoning whose conclusion is action.

*Trenton State College,*
*Hillwood Lakes, P.O. Box 940,*
*Trenton, NJ 98625, U.S.A.*

© GERALD W. BARNES 1983

---

# IS IT GOOD TO MAKE OTHERS DO SOMETHING GOOD?

*By* K. DOŠEN

IT doesn't seem wrong to assume that it is good to make others do something good. This assumption is certainly not foreign to the Christian tradition, with its insistence on spreading the Word. But the acts justified by this assumption need not be only verbal. For example, Saint Francis is reported to have said that it is good to beg because by begging others are induced to justify themselves by giving alms.[1]

However, this assumption can also lead to rather unacceptable conclusions. What would be thought of a man who went to drown himself in order to make others save him? Probably the most perverse use of this assumption was made by the gnostic sect of the Cainites who taught that Cain and all the worst characters of the Bible were the best. According to them Judas Iscariot did a good deed, since by betraying Jesus he made him redeem the world.[2]

---

[1] See Thomas de Celano, *Legenda secunda Beati Francisci*, 71, in *S. Francisci Assisiensis vita et miracula*, E. d'Alençon (ed.), Rome, 1906, and also *Speculum perfectionis*, ch. 18, in *Collection d'études et de documents sur l'histoire religieuse et littéraire du Moyen Age*, tome 1, P. Sabatier (ed.), Paris, 1898.

Let us briefly reconstruct an argument leading to a perverse conclusion on the line of those of the Cainites, in order to see that the problem is really with this assumption. Assume that x must choose between doing a good thing A and doing a bad thing B. If he chooses to do A, or B, y is bound to do the remaining thing. An example for such a situation can be invented with two knights having to decide who will take the sword to deliver a lady from a dragon, or something similar. Now, if we assume that

(I) *It is good to make others do something good*

and if x chooses to do B, he is doing something bad by doing B, but he is doing something good by making y do A.

If we want to get rid of such perverse conclusions, one possibility is to reject (I) outright. This rejection might be sustained by the following considerations. Doing A or B is doing something good or bad at the first level. Making y do A or B is doing something good or bad at the second level. Whereas it is our duty to be thoughtful of others and do only good things at the first level, at the second level we can (or perhaps even *should*) be selfish, and we are free to do something bad, if this means preventing others from doing good things at the first level by our doing these good things. Whereas it is good to yield one's place on earth, we should fight for our place in heaven.[3] So, doing something bad at the second level might be really good, and doing something good at the second level need not be really good.

What if we continued this ascent and introduced a third level where x does something good or bad if he makes y do something good or bad at the second level? And what if we continued this ascent still higher up? Should we do good things or bad things at these higher levels? Even if doing something good or bad at the second level makes sense, it is certainly strange to ascend much further. At those levels above the second which are still meaningful we should probably expect to find, *mutatis mutandis*, what we have found at the second level.

In general, the outright rejection of (I) doesn't seem very promising. Couldn't we say for a moral preacher or a legislator that the good he has done consisted exactly in his having induced others to do something good? But if we want to keep (I), it seems we must add some provisos in order to cope with perverse conclusions.

One possible proviso is to say that the good done by the act of making others do something good must outweigh possible bad aspects of this act. In order to guarantee that and not be entangled in a kind of moral calculus, we can reduce these bad aspects to nought, and simply assume that by making others do something good *we are not allowed to do something which is otherwise considered bad.* (In the terminology of levels we would say that by doing something good at the second level we are not allowed to do something which is bad at the first.) As a more explanatory premise from which this proviso can be derived, we can invoke the principle that we can't use bad means for good ends. We could also perhaps say that making others do something good is not intrinsically good, but only instrumentally good. Of course, the proviso above, or its weaker version involving a certain moral calculus, could be appended to other moral rules and need not be peculiar to (I). On the other hand, (I) may need provisos which are more peculiar to it, and we shall mention two of these.

---

[2] See J. W. C. Wand, *A History of the Early Church to A.D. 500*, London, 1937, p. 46.

[3] Cf. '... the kingdom of heaven suffereth violence, and the violent take it by force.' *Matthew* 11, 12.

What if doing B in our argument is not doing something bad, but rather doing something neutral? It could probably be objected that doing something neutral in a situation where we could do something good is doing something bad; so it is impossible that doing B is just neutral. But let us assume that it is possible that doing B is neutral whereas doing A is good. Then we may still feel that it is perverse to say that by avoiding doing A and by doing a neutral thing B, x is doing something good since he leaves to y to do A. The proviso above cannot cope with this situation. A further proviso is needed and perhaps we could again invoke a kind of 'second-level selfishness' of morality. *Whenever we can do something good, we should not make others do it before we have done it ourselves.*

Sometimes the perversity of arguments of the Cainitic type arises not from the fact that x is to choose between doing A and doing B, but from the fact that doing A presupposes that B is done. Now to cope with the case when doing B is bad, we can invoke our first proviso. The case when doing B is neutral need not seem perverse, and need not necessarily require a proviso to be blocked. But such cases are on the limit and we can still find in the vicinity goods being done by making others do something good which are rather spurious. For example, the good which we are making others do might be *artificial*, in the sense that the need for this good did not exist prior to our attempt to make others do it, and is provoked by this very attempt – very much as in the case of the man who goes to drown himself in order to make others save him. Saint Francis probably didn't want his brethren to be 'artificial' beggars, otherwise not needing alms. To conclude we can add a further proviso to (I) which should ensure that *we are not making others do artificial goods*, where 'artificial' is understood as above.

*Generala Ždanova 19,*
*11000 Beograd, Yugoslavia*

©K. Došen 1983

# NAGEL, WILLIAMS, AND MORAL LUCK

*By* JUDITH ANDRE

BERNARD WILLIAMS and Thomas Nagel begin their discussions of 'Moral Luck' by contrasting morality with luck.[1] Morality – at least as Kant articulates it – is the sphere of life in which, no matter what our circumstances, each of us can become worthy. Moreover, moral worth is the highest worth of all, and so there is a kind of ultimate justice in the world: each person is equally able to achieve that which matters most in life.

Nagel and Williams argue, however, that in practice we evaluate actions and agents partly on the basis of circumstances beyond the agent's control – on the basis of luck. Nagel argues that we ascribe *moral* value partly on such a basis, Williams that 'rational justification' rests partly on luck. For Nagel the consequence is an incoherence within our moral conceptual scheme. Williams reaches a similar conclusion more slowly. Since rational justification is partly a matter of luck, rational justification is not synonymous with moral justification. If this is so, then our notion of rational justification is not synonymous with that of moral justification, and morality is not the unique source of value; and if not unique – he argues – not supreme. 'If the moral were really supreme, it would have to be ubiquitous: like Spinoza's substance, if it were genuinely unconditioned, there would have to be nothing to condition it' (Williams, p. 38). But this is fundamentally different from the concept of morality we (in some sense) have now: 'one thing that is particularly important about ours is how important it is taken to be' (Williams, p. 39). For Williams, when luck enters into judgments of justification, the judgment is non-moral but competitive with – and so destructive of – moral justification. For Nagel, luck does enter into moral assessment, although our intuitions say it should not. Each writer finds destructive inconsistency – possibly incoherence – within our concept of morality.

To these charges, two responses are possible. One is that we do not in fact justify actions as it is claimed we do, or at least *would* not if we were sufficiently reflective. A second possible response is to admit that we do justify actions partially on the basis of luck, accept the conclusion that we are not consistent Kantians, but reject the implication that our moral scheme is therefore incoherent. I shall make both of these responses here.

I will discuss first the four areas where, Nagel claims, luck partially determines 'moral' value. One is constitutive luck: 'the kind of person you are ... your inclinations, capacities, and temperament'. Another is luck in consequences: 'luck in the way one's actions and projects turn out'. One drunken driver kills a child, but another gets home safely.

---

[1] Thomas Nagel, 'Moral Luck', in *Mortal Questions* (Cambridge: Cambridge University Press, 1979). Bernard Williams, 'Moral Luck', in *Moral Luck* (Cambridge: Cambridge University Press, 1981).

A third area is 'luck in how one is determined by antecedent circumstances'. This third area involves the classic conflict between believing we are determined and believing that we are morally responsible. So does Nagel's fourth area, which he calls 'luck in circumstance'. The person who became a Nazi prison guard might simply have been a bureaucrat, had his parents emigrated to Canada when he was young – yet we hold the guard morally responsible. This kind of example illuminates particularly well the difficulty in reconciling determinism and moral responsibility. Since we evaluate these two people differently on the basis of their different actions, we must be implicitly assuming some source of agency other than character (assumed to be identical in the two men) and circumstance (assumed to be beyond the control of either).

But compatibilism is not Nagel's primary subject. Since so much has been said on this issue already, and since Nagel does not advance that discussion, I would like to bracket the issue here.

The other areas, however – constitutive luck and luck in consequences – raise issues that can be discussed without entering the free-will-determinism debate. Take for instance the category 'luck in consequences'. Is the drunken driver who kills a child morally worse than the drunken driver who manages not to? Is successfully rescuing someone from a burning building morally better than trying unsuccessfully to rescue him? Or than accidentally harming him during the attempted rescue – perhaps by dropping him from a high ladder?

As I mentioned earlier, two kinds of response are possible. One is simply to deny the implication, to assert that these people do not differ in moral status. I'll make a more minimal claim here: the difference in moral evaluation substantially lessens upon reflection. The 'lucky' drunken driver has done something seriously wrong. Anyone who has lost a relative in an accident involving an 'unlucky' drunken driver will look with loathing upon *any* drunken driver. Reflection has the opposite effect upon our evaluations of lucky and unlucky careless drivers. If the carelessness is particularly common – taking one's eyes briefly off the road, for example – we're likely to say 'Anyone might have done that' and lighten our condemnation of the driver. In both cases, reflection at least narrows the apparent moral difference created by luck.

A variation of this response is to admit that luck does enter our moral evaluation, but to treat this as an inconsistency of which we could and should purge ourselves. Both Nagel and Williams claim that doing so would narrow the area of moral concern to the vanishing point (Williams, p. 38; Nagel, p. 35). If we were to exclude *everything* which Nagel calls luck – the kind of person one is, the circumstances which confront one, the choices one makes (if these are taken to be the result of antecedent circumstance), and the results of those choices – then Nagel and Williams are right. But in this paper I am bracketing the issue of determinism. Consistency in the remaining areas (constitutive luck and luck in consequences) would narrow but not erase the sphere of moral concern. Most people already make consistent efforts to separate moral judgment of someone from their appraisal of that person's 'constitution' (inclinations, capacities, temperament). Phrases such as 'Only God can judge her' reflect the belief that her internal intention to do the right thing – unknowable to the rest of us – is what really counts. As for luck in circumstances, it might be possible to learn to classify all careless drivers (lucky and unlucky) as morally identical.

However, a different kind of response is possible. Suppose that we do evaluate the lucky differently from the unlucky, and that we cannot (in some sense of 'cannot') change what we do. What follows from that fact? Although the remaining use of moral categories is clarly not Kantian, it does not follow that it is an inconsistent use. The word 'moral' is not nearly so precise as we sometimes feel – this is Nagel's and Williams' point – but it does not follow that the concept is internally incoherent.

The most obvious non-Kantian element in our moral scheme is the Aristotelian.[2] Part of being moral, for most of us, is being virtuous; and being virtuous involves more than doing the right thing. It involves as well the ability to see what the right thing to do is, and the *desire* to do that right thing. (Not just Kant's more abstract desire-to-do-whatever-the-right-thing-turns-out-to-be.) Virtues, as Aristotle describes them, are possible only to those who have been reared in a moral community; a fortunate childhood fosters adults who feel rightly as well as acting rightly. But people cannot choose their own upbringing, and emotions are not voluntary. (We do have indirect control over our emotions, as L. A. Kosman points out, but this is limited.[3] For Aristotle it is better not to have to struggle to do the right thing. For Kant the presence or absence of inner struggle is morally irrelevant. The word 'moral' has a somewhat different sense in the two writers, but there is this in common: for both the adjective refers to an excellence of character such that the moral person is praiseworthy and emulable. He or she is a model for our children, and the kind of person which we would like our communities to foster. The Kantian concept, however, unlike the Aristotelian, is closely linked to Christian ideas of reward and punishment by an all-just, omniscient Judge. Even an atheist can ask what such a Judge *would* do; and of course the

---

[2] Alastair MacIntyre, *After Virtue* (Notre Dame: University of Notre Dame Press, 1981). Although I use the terms 'Aristotelian', 'Kantian', and 'Christian' here, I am not defending any particular historical view about the content of these traditions. Instead I use the terms roughly, as convenient labels for identifiably separate strands within our present ways of thinking.

[3] L. A. Kosman, 'Being Properly Affected: Virtues and Feelings in Aristotle's Ethics', in *Essays on Aristotle's Ethics*, Amelie Oksenberg Rorty, ed. (Los Angeles: University of California Press, 1980).

Judgment could only concern those matters over which the agent had control. Reward and punishment could not justly be allotted on the basis of upbringing or genetic endowment.

We use the concept of moral in both the Aristotelian and the Kantian sense. Many of our central concepts are hybrid: we can call someone 'evil', for instance, without imputing responsibility (for she may be criminally insane). Someone may be sadistic because of the way he was raised – but he's still sadistic, and it is not a contradiction to say this is a moral fault for which he is not to blame. Open-mindedness comes more easily to some than to others; it's still (when properly limited) a moral asset. Similar remarks could be made about selfishness, generosity, cruelty, bravery – and even about the spoiling of children. Through no fault of their own, they've been made defective. Of course our condemnation increases as the child becomes adult – blameworthiness is a specific kind of criticism which implies free choice – but selfishness and petulance in children are nevertheless character defects.

Is a moral framework which includes both Kantian and Aristotelian elements self-contradictory? It can look that way, as Nagel and Williams' examples make clear; but the appearance results from assuming that 'moral' can only have a Kantian sense, and then discovering cases where its application depends on circumstances beyond the agent's control. In its central sense, I contend, morality refers to excellence of character. Whenever we praise people as moral we mean they are worthy of praise and emulation; but only sometimes do we mean that they are worthy of reward.

Let me apply this to Nagel's 'luck in consequences' – the cases of the lucky and the unlucky bad drivers, of the successful and unsuccessful rescue attempts. Some people persistently misjudge, and in the process hurt other people. (We might call the agents 'morally accident-prone.') They are malformed in some way; something prevents them from correctly assessing the facts before they act. Impulsiveness, or heroic fantasies, or self-absorption – whatever its source, their ineffectiveness make them less than admirable, less than a model. But to the extent that its source is beyond their reach, they are neither blameworthy nor punishable. This accounts for some of the 'moral distance' between our evaluations of the drivers, too; some people know their limitations better than other people do.

Furthermore, we say that the unlucky driver is morally worse than the lucky one in part because we hold people responsible for the results of their actions. But 'responsible' has two kinds of application. One is what I've called the Christian-Kantian sense: the responsible party is subject to reward and punishment. The other sense is more prosaic: to be responsible is to have an obligation to rectify bad consequences. If I break your vase, I must replace it. I can be responsible in the second sense without being in the least blameworthy, although often the two coincide. We can also be blameworthy (for the risk we take) without in fact bringing about any bad consequences, and so without being obliged to rectify anything.

At this point one may ask, obliged on pain of what? Certainly there is moral defect if we fail to make good what we have destroyed when it is possible to make this good; furthermore, the defect is greater when the source of the damage is not just our action but our culpable action. But what about the cases where the damage cannot be repaired? Loss of life, most obviously, cannot be put right. There is at least a great sadness in having incurred a debt which one cannot meet, and therefore a feeling of inadequacy. As a result of that there is a sense of diminished worth. This is, however, distinguishable from moral fault in the sense of deserving punishment. If these considerations are right, then part of our sense of 'moral distance' between the lucky and unlucky driver is neither derived from nor inconsistent with Kantianism.

These same considerations apply, more simply and directly, to what Nagel calls 'constitutive luck'. Some people are generous, brave, and honest by inclination; others are not. The fortunate ones are so constituted by upbringing and biology. They may deserve no particular reward for what they are; but they are nevertheless worthy of admiration and imitation.

I think, then, that Nagel's claim of contradiction within our moral concepts is mistaken. What about Williams' related claim that the justification of actions (as, roughly, the right thing to have done) depends partly on luck? If that's true, he claims, then rational justification is not synonymous with moral justification, and hence moral justification is neither the unique nor the supreme source of personal value. His argument is intricate, and I will address only a few strands of it here. He describes in particular the cases of Gauguin and of Anna Karenina. Each abandoned a family in order to pursue a future good (artistic accomplishment, on the one hand; passionate love on the other). Williams claims that these undertakings turn out to have been justified only if the goal is realized. This is close to the kind of situation Nagel describes as luck in consequences, and it is open, I think, to the same analysis. First, it's at least reasonable to claim that consequences have nothing to do with our judgment of the cases. Many would simply condemn both Gauguin and Karenina; others would approve their actions. Even if, as Williams suggests, we ask whether their actions were reasonable (rather than moral) it's quite possible to assess the reasonableness of risk-taking.

Finally, however, if we suppose that Gauguin's success does affect our evaluation of his choice, we can account for that on the Aristotelian grounds mentioned earlier. The person who can correctly assess his or her chances of success is better-formed than the person who cannot.

Since considerations of luck enter without paradox into one kind of moral evaluation, they do not indicate a non-moral, competitive realm of evaluation. It is still possible to characterize the morally good life as that which should be sought, all things considered.

A final note. Not only does 'moral luck' pose considerably smaller problems than Nagel and Williams suggest, it is in fact part of our conceptual scheme which should, upon reflection, be kept rather than changed. First of all, it reminds us of moral questions different from those upon which we typically concentrate. 'What should I do in these circumstances?' is important; but 'What kind of person should I try to be, and help others to be?' is of equal importance. Secondly, 'moral luck' is sometimes illusory; we are sometimes to blame for results that at first *seem* beyond our control. We have more control over the kind of person we are than we sometimes think. I can, for instance, whittle away at the habits of self-deception which prevent me from seeing the world as it is. The 'morally accident-prone' can learn to be better; but they are unlikely to change if all our moral assessment is concentrated on intention and none on actual result.[4]

*Old Dominion University,*
*Norfolk, Virginia 23508, U.S.A.*

© JUDITH ANDRE 1983

---

[4] This paper received the 1983 Griffith Memorial Award from the Southern Society for Philosophy and Psychology, U.S.A.

# THE TOXIN, THE BLOOD DONOR AND THE BOMB

By JANICE THOMAS

WHEN I form an intention to $\phi$ it sometimes happens that my formation of the intention has consequences independent of the effects, if any, of my $\phi$-ing. My present intention to give blood tomorrow might inspire a new blood donor today – whether or not I go through with it. Thus, knowing that my terror of the needle makes me a persistent akratic where giving blood is concerned, I might still operate an effective scheme to recruit donors by manifesting a good strong intention to donate (and disguising my full awareness of my past weaknesses) on suitable occasions.

Now Gregory Kavka ('The Toxin Puzzle', ANALYSIS 43.1, January 1983, pp. 33-6) might concede that I could inspire with my beneficent intention-forming, but would be quick to distinguish my akratic blood donor story from his tale of the toxin. For his protagonist is faced with the challenge to gain the fruitful consequences which would flow from his forming a simple intention when he has no reason to do the action intended and good reason not to. To gain the eccentric billionaire's million dollars all the protagonist has to do is form the *intention* to take the toxin –

thereafter he will receive the money, whether he carries out his resolve or not. So he has no reason to take the toxin and good reason, in the day's illness it would cost him, not to take it.

Kavka suggests the protagonist cannot get the money because he cannot intend just whatever he wants to intend any more than I can believe just anything I wish to believe. This being so, however, it seems that Kavka cannot generate the paradox he hopes to generate in a third sort of case – the case of the would-be deterrer of nuclear attack (see his article 'Some Paradoxes of Deterrence', *Journal of Philosophy* LXXV 6, June 1978). In that example the would-be deterrer thinks that being perceived genuinely to intend to resort to nuclear retaliation in the event of nuclear attack might well be sufficient and (bluffing being too risky) necessary to deter such attack. His problem is that he knows he would not resort to nuclear retaliation if his nation were devastated by nuclear attack since such retaliation would serve no defensible purpose and would involve injustice and disutility of extraordinary proportions, thus being morally abhorrent. The supposed paradox is this. It would be right for the agent to form the intention to retaliate since that genuine intention perceived as such would prevent nuclear attack. However, the agent, knowing he would never in fact retaliate ($\phi$), cannot form the desired intention. He would have no reason to $\phi$, and the strongest reasons not to $\phi$ should the circumstances for $\phi$-ing arise. So he cannot now do what is right: only a less moral agent (who thought retaliation not under a moral ban) could do so.

Certainly there is an air of paradox here but the toxin puzzle provides not just another example of the same paradox but the wherewithal to dispel it. If I cannot form the intention to retaliate then it is erroneous or disingenuous to judge that I *ought* to form that intention. Whatever clout is left in the slogan 'ought implies can', surely it can be wielded here. I cannot have a duty to do the impossible. The appearance of paradox comes from an equivocation on 'right' (and 'ought'). Possibly forming the intention to retaliate would be right in the sense that good deterrent consequences might flow from that intention if, *per impossibile*, it could be formed. But no moral system, whether act- or agent-oriented, could conclude that I ought (i.e. have a duty) to form that intention where forming it is impossible.

But, it will be objected, nuclear deterrence is not a logical impossibility. Surely it is perfectly possible to form a conditional intention to retaliate hoping thus to forestall any chance of the condition's being met? On the contrary Kavka has shown that this apparent possibility is merely that – apparent but unreal. That deterrence is none the less possible is a function of the fact that although we cannot form the intention of counterattacking where we would not do so, we certainly can do some things that look efficaciously like it – threaten to strike back if struck; issue standing orders to retaliate not intending them to be obeyed; tolerate nuclear weapons on our soil albeit having no intention of using them and foist such weapons on our friendly neighbours and allies.

It is a mistake to regard a nation as exactly analogous to an individual moral agent in forming, publicising (or hiding) and acting on intentions. When *I* have (or lack or cannot have) a particular intention this is a single fact about me which I suspect it would be laughably easy to discover. In contrast, the intentions of a nation are abstracts – based on but not easily identified with the intentions of the political leaders and the military and also the electorate, at any rate those who hold the majority view. (And we might ask how much weight, if any, an enemy nation trying to assess the threat of counterattack by us ought to give to opinions expressed in our media.) Even if an implausibly extensive spy network showed that we were all, to a man, incapable of forming the intention to counterattack, I doubt whether our enemy's leaders would or should feel sanguine about their chances of making a nuclear attack with impunity. They would be fools if they did not give ample weight to the reflection (familiar to even the most enlightened of parents) that a punitive retaliatory act is astoundingly easy to provoke even from the hand of one who had never intended to retaliate. Moreover, if I am right that nuclear deterrence does not require the formation by any agent of a genuine intention to retaliate then clearly we should be deeply sceptical about the suggestion that a moral agent might ever have a duty to corrupt himself in order to be able to form such an intention for purposes of deterrence.

To return to the toxin puzzle: if I were some day lucky enough to be offered the eccentric billionaire's challenge there are two directions in which I would seek surety of gaining the million dollar prize. First, I would remember the akratic blood donor. My desire for the money might not give me any reason to take the toxin but it would give me excellent reason to search my personal history, present circumstances and character for *other* reasons I already have for taking the toxin. The akratic has reasons for doing what he intends but his incentive to be guided by them is badly eroded: in contrast, I would have ample incentive to be guided by whatever reasons there were for me to take the toxin. Nor would searching for such reasons break the contractual prohibition against the 'arrangement of external incentives' and 'alternative gimmicks'. I would not be importing anything external but employing reasons (albeit perhaps frivolous) I already possess.

And of course I could find several such reasons if need arose. (Is it ever true, about *any* action I consider doing, that I have *no* reason, however insubstantial, for doing it?) In my case appetite for personal celebrity, greed and curiosity give me plenty of scope. The newspapers would be far more eager for my story if they could print both a picture of me swallowing the mystery substance and my graphic account of my subsequent (if brief) agonies. They might even pay – I quite relish the prospect of further profits piled atop my million. Anyway I really would like to know what the stuff tastes like. Where there is a will this is often the way – a piece of bigotry to be buttressed, a self-serving belief to be protected, and we find ourselves able to tinker with our belief picture, altering the composition and the depth of field, bringing certain evidence into sharp focus while blurring and playing down other considerations. Self-deception, hypocrisy, and the forming of intentions on inadequate grounds are all of a piece and all very live options.

Meanwhile, I am doing my best to remove the one reason I would have not to take the toxin. I have got my husband – the crack biochemist – hard at work on an antitoxin. He expects a breakthrough any time. I shall have no trouble forming my good intention when the eccentric billionaire knocks.

*Heythrop College,*
*11–13 Cavendish Square,*
*London W1M 0AN*

© Janice Thomas 1983

---

# HUMAN REPRODUCTION AND SOCIOBIOLOGY

*By* John Dupré

LYNNE M. BROUGHTON's paper 'Human Reproduction: A Self-defeating Strategy' (ANALYSIS, 43.1, January 1983, pp. 54–8) purports to show that the theory of human sexuality suggested by sociobiologists is incoherent. However, the account she gives of this theory is an extremely lopsided one, and the argument she offers against the part of the theory that she does emphasize is unconvincing.

The particular thesis that Broughton singles out for attack is the prediction that an important part of the female reproductive strategy will be the insistence on finding a male who will cooperate in the rearing of offspring. As a matter of fact I do not think this thesis is especially plausible, and I shall explain very briefly why this is so at the end of the paper. However, what I wish to say in reply to Broughton is first, that this thesis is, contrary to the impression she gives, quite peripheral to the main thrust of sociobiological theory;[1] and second, that the arguments she offers are misguided, and importantly so.

---

[1] Broughton suggests that Wilson considers this an integral part of the theory by selective omission from a quotation. She quotes him as writing: 'In theory it is more profitable for females to be coy, to hold back ... to select males who are more likely to stay with them after insemination'. The omission is: 'until they can identify males with the best genes. In species that rear young, it is also important for the females to select', etc. (E. O. Wilson, *On Human Nature*, Cambridge, Mass., 1978, pp. 124–5). The omitted part is the crux of the theory, the thesis Broughton attacks a far more debatable addendum.

To begin with the first point, what is typically emphasized by sociobiologists is that each sex can be expected to pursue a reproductive strategy that will maximize the genetic contribution of successful individuals to the next generation. In the case of males, for whom reproductive potential is virtually unlimited, the strategy predicted is promiscuity. Females, on the other hand, can typically only hope to rear a more limited number of offspring, and their optimal strategy will be to maximize the chances of survival for the relatively few they can produce. One obvious way of doing this is to be as selective as possible about the male with whose genes theirs must coexist. This insight seems to me to be highly illuminating, and borne out by a great deal of ethological data.

The *universal* reason why females have a more limited reproductive potential is that their physical contribution to the offspring, in whatever form it is produced, is vastly greater. In a relatively small number of species the offspring require considerable postnatal rearing, and this raises a further issue, the one with which Broughton's paper is concerned. It is obvious, from an evolutionary standpoint, that each sex would be better off leaving this task to the other. It is also fairly obvious that the female is in a weaker position in this respect. The male is typically able to take off after fertilization, and leave the female, literally, holding the baby. In the majority of such species this is exactly what happens though, quite surprisingly, sometimes the male does the work (e.g. in Darwin's frog, the male of which species raises offspring in his voice pouch). It is very plausible that if there were some strategy by which the female could induce the male to help with the work of rearing, she would attempt to deploy it, and the males would attempt to resist it.

Broughton argues that such a conflict is impossible, since if such strategies were implemented the females would only accept a mate who demonstrated his willingness to stay around, and no such males would be forthcoming. Thus no reproduction could take place. One might just as well argue that since it is impossible for both players to win at chess, no one could ever play. Broughton does consider the possibility that some compromise might be reached, but dismisses it on the ground that no set of strategies that required such compromise could be optimal for the species, and hence, apparently, it could not be actual. But it is quite generally unsound to argue that an optimal solution will always be arrived at by evolution. (Why don't snails have wheels?) Evolutionarily stable strategies, to borrow an extremely illuminating concept from sociobiology, are by no means typically optimal for the species.[2]

---

[2] A more profound objection is to the whole notion of evolutionary advantages being relevant to *species*. A major thesis of the work by Richard Dawkins that Broughton cites (*The Selfish Gene*, Oxford, 1976) is that evolution should always be understood in terms of reproductive advantages to *genes*. Her extraordinary attribution to Dawkins of the thesis that genes have 'an urge ... to perpetuate themselves' does not suggest a profound understanding of this work.

Broughton continues by observing that sexual behaviour in animals is 'highly structured and complementary' and hence shows no evidence of conflict. This evinces a fundamental misunderstanding of the role played by the notion of conflict in evolutionary theory. The evident existence of conflicting reproductive interests is used to explain the evolution of behaviour patterns that are optimal for both parties given the strategy evolved by the other side. This is what an evolutionarily stable strategy is. Clearly this precisely predicts 'structured and complementary behaviour'; but far from being incompatible with antecedent conflict, such conflict explains the structured behaviour. (It is worth adding here, that sociobiology does indeed predict *overt* conflict over sex, though not between the sexes, but between the males. This prediction is amply borne out.)

The real trouble with the thesis that females will attempt to insure male cooperation in child-rearing is that it is hard to see what strategy could possibly achieve this unless *his* reproductive interests were also served. This would clearly be the case if two parents were required for rearing the offspring (or at least the offspring would have a greatly increased chance of survival with two cooperating parents). Presumably this is the case for many birds, where one parent must look after the nest while the other forages. (This is also the case in which one might predict monogamy.) But neither observation of humans, nor of their closest relatives, suggests that anything of this kind is true for them. Hence there is little reason to suppose that evolution will tend to lead human males to cooperate in child-rearing or human females to coerce them into doing so. Indeed, since divorce proceedings generally concern the *right* to bring up children, rather than the *duty*, there is good reason to conclude that this aspect of human behaviour is very largely determined by culture. But this by no means implies that sociobiology has no insights to offer us about sexuality, human or otherwise.

*Stanford University,*
*Stanford, California 94305, U.S.A.*

© JOHN DUPRÉ 1983

---

# THE COHERENCE OF THE DOUBLE STANDARD

*By* JOHN C. BIGELOW *and* ELIZABETH W. PRIOR

IT has been argued (e.g. E. O. Wilson, *On Human Nature* Cambridge, Massachusetts, 1978; R. Dawkins, *The Selfish Gene* Oxford, 1976) that there is a biological basis for the double-standard within human sexual morality. In 'Human Reproduction: A Self-Defeating Strategy' (ANALYSIS 43.1, January 1983, pp. 54–8)

Lynne Broughton argues that any theory which attributes conflicting reproductive interests to males and females 'is incoherent because it involves a set of reproductive strategies which for the species as a whole would be self defeating' (p. 54).

The first thing that we need to get clear is exactly what behavioural traits we are talking about. In the case of Dawkins one important behaviour differential (purported behaviour differential) he is interested in providing an account of is that women *themselves* prefer to be sexually monogamous, and men *themselves* to be polygamous. It is clear that such a difference in behaviour need not necessarily generate conflict. A woman can be totally monogamous and her one sexual partner radically promiscuous, and both could be quite content with that arrangement. Of course, this would entail that some males are left without any sexual partners at all, given roughly equal numbers of males and females. Broughton mentions the case of walruses, as one in which 'all the females mate with one or a number of dominant males and weaker males lose out' (p. 57). Such a situation need involve no *conflict* in the reproductive interests of males and females: though of course there will be a strong conflict in the interests of successful and as-yet-unsuccessful males. In addition to the question of how the sexes themselves prefer to behave there is also the question of how an individual would like his/her sexual partner(s) to behave. A promiscuous man may (and frequently does) have a strong preference for his sexual partners to be faithful to him, even while himself having no inclination toward, nor intention of observing such behaviour. Broughton's discussion indicates that when she talks about the female preference for monogamy what she has in mind is a preference for the monogamous relationship, that is the monogamous behaviour of *both* parties.

The conflict that here arises is a conflict concerning the male's behaviour: the male prefers promiscuity, the female that her mate be faithful to her. Broughton argues that in a species composed of such individuals, if those individuals acted always and only in ways the theory claims are their best reproductive strategies 'it would follow that, barring rape, no women would become impregnated since they would be waiting for men who would stay with them – yet there would be no such men' (p. 56).

## I WOULD A DOUBLE STANDARD CAUSE EXTINCTION?

Broughton's argument cannot be one against conflict models *per se* because there are numerous instances where an entity benefits in virtue of the simultaneous operation of processes with conflicting ends. Within the cell for example there are processes (anabolic) which convert sugars to animal proteins and processes (catabolic) which convert animal proteins back to sugars. The continuous simultaneous operations of both sets of processes is necessary for the well being of the cell, and the organism composed of such cells. In a similar way, an economy can operate very productively, even though it contains rival companies competing for the same consumers, and even though there is conflict between the interests of employers and employees. The mathematical development of games theory, on which Dawkins, for instance, rests much weight, has demonstrated a variety of ways in which a state of equilibrium can result from the interaction of parties with conflicting interests.

So why does Broughton consider a theory which attributes conflicting reproductive interests to males and females incoherent? In the passage quoted above she argues that in a species composed of promiscuous males and females who desire monogamous relationships, if those individuals acted always and only in the ways the sociobiologists claim are their best reproductive strategies, no intercourse would occur and thus there would be no reproduction.

The sociobiologist (Dawkins for example) explains the female's preference for a monogamous relationship in terms of it being *a* strategy which endows her genes with a better chance of getting into succeeding generations. And, presumably, there will be *other* strategies 'aimed' at the achievement of the same end. The females preference for physically robust males with power and resources is a preference for which Dawkins can and does provide the same explanation. Alliance with such males will bestow upon the female's genes a better chance of getting into succeeding generations.

But then there is little reason to expect that the same individuals must always satisfy both these female preferences. There is no reason to suppose that monogamous males will always be the most physically robust and the ones with the greatest power and resources. Indeed one would expect the opposite would often be the case.

Frequently a strong, successful male, because he has more resources at his disposal, will be able to provide more for each of a number of females than she could get if she allied herself with a faithful but less successful male. In such a case the female will have a better chance of perpetuating her genes if she allies herself with the non-monogamous male.

The females preference for a monogamous relationship is simply *one* strategy for propagating her genes. The sociobiologists named by Broughton never suggest it is the only strategy.

The second criticism which can be made of Broughton concerns her 'barring rape' rider in the passage quoted. It is fashionable nowadays to view the natural as the good and the bestower of happiness. This is expressed in the commonly held belief that if we were to remove humans from this grasping, competitive, sexist culture they would then treat each other with kindness, tolerance and concern. But there is virtually no psychological or anthropological evidence to support this halcyon vision. When human beings wandered the Savannah in tiny groups sexual congress need not have been a meaningful experience between mutually consenting parties. On the contrary it may well have been paradigmatically nasty, brutish and short, at least for the females – something frequently approximating what we would rightly call rape.

If a male perceived a sexually titillating female tripping alone across the grassland then the normal course of events would be for him to have intercourse with her, whatever her preferences might be. The procreative act requires virtually no female co-operation. Whether she consents and enjoys it or whether she resists till the last does not affect whether she is impregnated on that intercourse. On the other hand, in order to avoid tripping alone across the grasslands, a female might find it better to seek the protection of a powerful male. And this would entail that she submit to his sexual advances, whether or not his behaviour in other respects was as she might wish it to be.

Broughton suggests that at the relevant period of human evolutionaty history human females had the option of refusing intercourse and thus impregnation (see the passage quote above). But those females had no such option. One way or another the chances of a female being impregnated were high. To the extent that she had any choice therefore it would be in her own and her genes' best interests to ally herself with the best of the available males (imperfect though they may have been).

To sum up, sensible sociobiologists do not suggest that females follow the injunction: 'Mate only with faithful males'. Rather, they suggest that females follow some injunction more like this: 'If you have a choice, then, other things being equal, choose the *more* faithful male'. Similarly, sociobiologists need not postulate a conflict of interests so gross that males follow a totally undiscriminating policy of 'casting large numbers of offspring upon an uncaring world' (p. 57). They may only be following the more subtle strategy of producing offspring from as many females as they can support themselves, *and* as many as they can find who can support themselves or be supported by others. Again, such a more discriminating policy complicates the picture, but does not remove sexual conflict of interests altogether.

## II WOULD A DOUBLE STANDARD BE UNSTABLE?

Let us return to Broughton's central claim: that the existence of conflicting reproductive interests for males and females is an 'incoherent' supposition, since 'it involves a set of reproductive strategies which for the human species as a whole would be self-defeating'. The question arises just what she means here by 'self-defeating'. Sometimes she argues as though she means' 'causing extinction'. It is, indeed, 'incoherent' to suppose a species to have a characteristic which would cause extinction, and yet to show no signs of decreasing in numbers. But, as we have argued, the existence of a conflict in preferences between males and females need not result in failure to mate, and hence need not lead to extinction.

But perhaps a strategy could be 'self-defeating' in a less direct manner. For instance, on pp. 56–7 Broughton does consider the possibility that conflicting preferences could be over-ridden by 'negotiation', or by 'a range of factors which might minimize its effects'. She asserts that the fact that negotiation is *needed* shows the original strategy is 'unsatisfactory'. Another word she uses is 'optimal' – the existence of conflicting preferences is 'hardly optimal'. One may ask – unsatisfactory for whom? The walruses' reproductive strategy is unsatisfactory for most males, but that is not enough to prevent it persisting.

Suppose, however, that we ask whether the existence of conflicting preferences is optimal for the species as a whole. This may suggest the following line of argument. Suppose there to be competition for some ecological niche – competition between several species, or subspecies, or even amongst some smaller groups within a species. And suppose one such group was afflicted with a *greater* degree of conflict between the preferences of males, and of females. Then surely, it may be argued, the group with less internal conflict would be more successful. And so internal conflict within the group would be 'self-defeating', in the sense that it would result in a less successful group, which would lose out in the process of natural selection.

This argument is more worrying, since it has much weaker premises. It might be resisted, by Dawkins for instance, on the ground that there is something incoherent in any explanation appealing solely to what is optimal for the species *as a whole*. The unit of selection, says Dawkins, is the gene, not the species.

Fundamentally, the objection is this. The mechanism by which one generation of a species produces the next generation is by way of the transmission of genes by individuals. Suppose some hereditable characteristic benefits the species as a whole, but does *not* give any *greater* benefit to those individuals bearing the characteristic. Then that characteristic will not increase in frequency. A characteristic can only increase in frequency when it bestows a greater benefit on those who bear or share the characteristic. And conversely, a characteristic will decrease in frequency only when it confers a differential disadvantage on those who bear or share it. In other words, not only must it be less than optimal for the species as a whole, but also, it must be *more* disadvantageous to those with it than to those without.

Thus, even if a sexual double standard were less than optimal for the species as a whole, once begun it would nevertheless persist, provided the *first* 'reformers' were always at a disadvantage among a community of 'conservatives'.

Finally, there is yet a further reply possible on behalf of the double standard. It is not clear that a sexual double standard *would* reduce the reproductive rate for the species as a whole. Females might follow the injunction: 'If you have a choice, choose the more faithful male'. And males might follow the injunction: 'If you must, then be faithful, but cheat if you can'. That would surely count as generating a conflict of interests. And yet there is no reason why a species full of individuals following those preferences need produce any *fewer* offspring than a species with more harmony between the sexes.

The existence of a double standard may be unpalatable and unproved, but it is not 'incoherent'.

*La Trobe University, Bundoora,*
*Victoria, Australia 3083;*
*Monash University, Clayton,*
*Victoria, Australia 3168*

© JOHN C. BIGELOW *and*
ELIZABETH W. PRIOR 1983

# WHAT WAS HUME'S WORRY ABOUT PERSONAL IDENTITY?

*By* GEOFFREY SCARRE

'WHEN I turn my reflexion on myself', wrote Hume in the Appendix to the *Treatise of Human Nature* (ed. L. A. Selby-Bigge, Oxford: Clarendon Press, 1888, p. 634), 'I never can perceive this *self* without some one or more perceptions; nor can I ever perceive any thing but the perceptions. 'Tis the composition of these, therefore, which forms the self.' Hume was recalling the view he had confidently defended earlier in the *Treatise* (at I.iv.6), but his confidence had now evaporated. Hume's position in the Appendix was that there still seemed to be some force in the reasons for thinking of the self as no more than a time-extended bundle of perceptions related by resemblance and causation, yet he now felt the theory to be for some reason unsatisfactory: 'when I proceed to explain the principle of connexion', he confessed, 'which binds [particular perceptions] together, and makes us attribute to them a real simplicity and identity; I am sensible that my account is very defective, and that nothing but the seeming evidence of the precedent reasonings cou'd have induc'd me to receive it' (635). He professed to find the whole matter a 'labyrinth', and declared that he knew neither how to correct his former opinions nor how they could be rendered consistent.

Every writer on philosophical psychology has his pet theory of what is wrong with Hume's bundle conception of the self. My intention in this note is not to rehearse familiar criticisms, but to ask the question: what was *Hume's own* ground of objection to the bundle theory? Students of Hume are well aware that it is highly obscure just what complaint Hume was making in the Appendix. Of the several interpretations proposed, none, I think, has come very close to Hume's meaning, although some have raised points of considerable philosophical interest. (For a survey of recent interpretations, see Don Garrett, 'Hume's Self-Doubts About Personal Identity', *Philosophical Review*, XC, 1981.) I offer a fresh view of Hume's difficulty which I believe is better supported by the text than are those previously proposed, and show how Hume spotted a powerful objection to the bundle theory which has not figured among standard criticisms.

What Hume did make plain was that he thought his views about the self involved an inconsistency. Unfortunately he failed to articulate this inconsistency clearly. To be sure, he declared that 'there are two principles, which I cannot render consistent; nor is it in my power to renounce either of them, viz. *that all our distinct perceptions are distinct existences*, and *that the mind never perceives any real connexion among distinct existences*' (636; Hume's emphases). But, as all the commentators have noted, these principles are transparently *not* inconsistent. I agree with Barry Stroud that we should presumably take Hume as having meant that if we keep hold of these principles we somehow make it impossible for ourselves to explain the origin of the idea of personal identity (*Hume*, London: R.K.P., 1977, p. 128); but we need to mine to discover why he should have thought this.

In disentangling Hume's view we might usefully begin by observing that both the principles to which he drew special attention concern the notion of distinct existences. Moreover, Hume spent no fewer than three paragraphs of his Appendix note explaining how the objects which we distinguish in our perceptions are taken by us to form a manifold of separately existing items (634). This should alert us to the likelihood that the inconsistency troubling Hume resides somehow in the contrast between our standard ways of conceiving the objects of our perceptions, on the one hand, and ourselves, the percipients, on the other. Hume wrote about the objects of perception: 'All perceptions are distinct. They are, therefore, distinguishable, and separable, and may exist separately, without any contradiction or absurdity. . . . [T]his table, which is present to me, and that chimney, may and do exist separately. . . . But 'tis intelligible and consistent to say, that objects exist distinct and independent, without any common *simple* substance or subject of inhesion' (ibid.).

Hume's reflections on our manner of counting the objects of perception may not seem specially striking, yet they are of great significance in the context of a discussion of personal identity. If I look first at a table and then at a chimney, I do not, as Hume remarked, consider the table and the chimney to be *one* object; rather, I believe that my successive experiences have presented me with a plurality of objects, of separate existences – there are *two* things I see, a table and a chimney. Now a perception can be considered from two different angles, which might be termed the objective and the subjective: objectively, a perception may be *of a table* or *of a chimney*; subjectively, it might be *Hume's* perception of a table or *mine* of a chimney. On my reading of Hume, he believed he saw an incompatibility between regarding a sequence of perceptions as presenting, from the objective angle, a plurality of distinct existences, while yet constituting, from the subjective point of view, an unitary self. The same series of perceptions, Hume came to think, could not be said to present a number of non-identical objects (e.g. a table, followed by a chimney, followed by . . . etc.) and *also* be said to form a single self (or a segment thereof). On his view, 'no proposition can be intelligible or consistent with regard to objects, which is not so with regard to perceptions. But 'tis intelligible and consistent to say, that objects exist distinct and independent, without any common *simple* substance or subject of inhesion. This proposition, therefore, can never be absurd with regard to our perceptions' (ibid.). Our practice of regarding perceptions as presenting us with non-identical items provides, in Hume's eyes, a ground for refusing to look on series of perceptions as constituting single selves.

But why should such a line of thought have impressed Hume? There is certainly no logical contradiction incurred in holding a sequence of perceptions to be manifold from the objective point of view and unitary from the subjective. Hume presumably thought that some kind of explanatory tension was involved in bundling together to form the history of a single self a sequence of perceptions of a plurality of objects. But is there truly a tension here – particularly if the idea of an unitary self is taken, as Hume took it, to be a kind of convenient fiction anyway?

A bit of reflection should show that Hume's worry was not a negligible one. Recall the view of the self which Hume began by rejecting, namely that the self is an abiding immaterial substance capable of perceptions, a real thing and ground in which mental states inhere. Having thrown out this metaphysical self, Hume aimed to preserve some content to the notion of a persistent self by identifying it with a sequence of perceptions; when he spoke of selves as fictions, what he was telling us, in a slightly misleading way, was that the reality of selves was no more (but also no less) than the reality of those sequences: nothing substantial remains constant over time underlying perceptions as their true subject. Yet, as Hume willingly conceded, the revised view of the self remains thin on content until something is said about the manner of linkage of a series of perceptions. According to Hume, the agencies of unification of sequences of perceptions are relations of resemblance and causation among individual perceptions. A sketchy account of the importance of these relations for personal identity was provided at *Treatise* I.iv.6: the production in memory of images (counted by Hume among 'perceptions') resembling earlier perceptions helps to create the sense of a continuing self, an effect completed by the fact that perceptions are 'link'd together by the relation of cause and effect, and mutually produce, destroy, influence, and modify each other'; impressions give rise to ideas, which in turn produce other impressions, and '[o]ne thought chaces another, and draws after it a third, by which it is expell'd in its turn' (261). It is only because perceptions come in sequences linked by resemblance and causation that there is any point to talking of a persisting self; without these relations, perceptions would remain starkly individual and unconnected, making it quite impossible for the concept of personal identity to get any grip.

In the Appendix, Hume had to admit the inadequacy of this account. The argument which impressed him so much was an analogical one. We want to maintain that our perceptions present a manifold of objects, and not a single object. Yet – and here is where Hume has confused commentators by being less than fully explicit about how the argument goes – we perceive those objects as existing in causally patterned relationships with each other, and not as a confused jumble of unconnected particulars; resemblances and constant conjunctions are indeed among the most basic and familiar features of our experience. But although we find resemblances and causal connections among the objects of our perceptions, we are not inclined to bundle those objects together into one: we still regard ourselves as experiencing *a plurality of objects, and not a single object*. However, if resemblance and causation are insufficient in the case of objects perceived to import unity, then it cannot reasonably be held that these relations are sufficient to warrant talk of sequences of perception as constituting unified subjects. (This thought probably struck Hume the more strongly because of his adherence to the doctrine that causal relations are nothing more than constant conjunctions, and not 'real' connections of things.) The problem of accounting for our idea of personal identity is not, of course, hereby proved to be insoluble. But its solution, in Hume's eyes, became a matter of discovering some other agencies of unity which fare better than resemblance and causation, and he lamented: 'I cannot discover any theory which gives me satisfaction on this head' (636); once we have 'loosen'd all our particular perceptions' [i.e. from an underlying substantial self], and have identified no other relations than those of resemblance and causation as holding among them, then some attention to our manner of conceptualizing the objects of perception, which are also related by resemblance and causation, should convince us that we can no longer attribute to them any genuine 'simplicity and unity' (635).

In my opinion, Hume's despair about the capacity of the theory of I.iv.6 to explain the origin of our idea of personal identity is justified by the argument I have attempted to reconstruct. On the exiguous basis provided by that theory, the notion of a persisting self cannot even appear a persuasive and convenient fiction – resemblance and causation among perceptions are simply inadequate to sustain the content of the supposition of personal identity. Some may think to counter Hume's argument by urging that while we take the objects of our perceptions to form a plurality, they are nevertheless united in belonging to the *same world*, and that it is the world, rather than the objects it contains (or, variantly: which are bundled together to constitute it) which is the true objective analogue of the subjective self. But though Hume did not consider this counter-argument, it is very doubtful whether that omission seriously weakens his case. For the world is a very poor candidate for a simple and unitary object, and even someone willing to settle for a fairly austere conception of the self would be unlikely to feel that a comparison of the self to the world provided a remotely adequate vindication of the idea of the self as truly unified; for it is natural and reasonable to think that whatever relations unite the self (if, indeed, any do) must be of a more intimate kind than those which obtain among the objects which make up the world. Thus either the notion of personal identity is a complete illusion, or it needs to be accounted for in some way radically different from Hume's bundle theory.

*University of Durham,*
*50 Old Elvet, Durham DH1 3HN*

© GEOFFREY SCARRE 1983

---

# CONJECTURES AND MODALITIES

*By* BERNARD MAYO

I SHALL argue that George Mavrodes ('Vestigial Modalities', ANALYSIS 43.2, March 1983, pp. 91–4) has misidentified the origin of a very common type of modal function. (An incidental comment is that he seems unaware how very common the type is.) He claims that a statement such as[1]

(M2) She got married last week, so her father cannot do anything now to prevent her marriage

is, despite appearances to the contrary, 'quite neutral with respect to the question whether anything can be done now to prevent a past event' (94). For, he conjectures, the force of the modal term 'cannot' is to 'record and recall' a modality contained in a suppressed premiss of the inference of which the second component of (M2), is the conclusion, this component itself being properly construed as an unmodalized proposition:

(M12) Her father will not now do anything to prevent her marriage.

---

[1] Mavrodes' numbered propositions '(2)' etc. are repeated here as '(M2)' etc.

This conjecture is partly right and partly wrong. It is (partly) true of very many expressions like (M2), but it is not true of (M2) itself. The reason why it is only partly true even of the other expression-types is that, while the modal term does indeed reflect a context of inference from premisses, including suppressed ones, to an unmodalized conclusion, the modality is not, *contra* Mavrodes, a vestige or shadow of a substantial impossibility lurking in a suppressed premiss. Rather, it is a reflection of the inference itself, a modal term being a natural way of reflecting argumentative cogency.[2]

To see this, consider Mavrodes' prime example for persuading us of what he says about (M2):

(M6) The County Clerk has no record of them, so they could not have been married here.

He first suggests that we might interpret the 'could not' in terms of a *barrier* (racial prejudice, age limit, etc.), but points out, correctly, that there might be no such barrier yet the 'could not' might still be in order. His conjectural explanation of the apparent anomaly is that the second component of (M6) is really

(M11) They were not married here

which is to be understood as an inference from the following two propositions:

(M8) It is not possible that (both) the County Clerk has no record of them and they were married here

(and this is described as a substantial impossibility 'growing out of laws, bureaucratic practices, etc.'), and

(M9) The County Clerk has no record of them.

I want to question two things in this account: (i) the introduction of 'barriers', and (ii) the ascription of the 'shadow' modality of (M6) to the 'substance' modality of (M8).

(i) (This is where Mavrodes seems not to realize how very commonplace his example is.) Suppose a more successful family history researcher had actually found the record (of a marriage), and reported, instead of (M6),

(16) The County Clerk has a record of them, so they must have been married here.

Would Mavrodes have us start by imagining irresistible forces in place of unscalable barriers? A shotgun wedding? Of course not.

---

[2] I have developed this point at greater length in a paper 'On the Lehrer–Taylor Analysis of 'Can'-Statements', *Mind*, LXXVII, 1968, 271–8. I should not now be prepared to defend an implication of what I then wrote, that *all* uses of 'cannot' reflect argumentative cogency. In the present note I assume that there are *de re* necessities and impossibilities.

There *could* have been a report concluding with 'So they had to get married', but premisses about marriage records would not figure in it, either overtly or covertly. Instead, the very fact that we know what sort of premisses we do have, plus a very clear linguistic distinction between the different modal forms 'must have done' and 'necessarily did', points us in the direction of the truth that the function of 'must have' is to reflect conclusory force.

Mavrodes appears to have been misled by the fact that the obvious difference between 'must have' and 'had to' is not quite so clearly repeated in the corresponding negative forms which occur in his examples. And 'he could not have done' is indeed ambiguous as between 'it cannot be the case that he did' and 'he was unable to', but these are clearly different. Note that (M6) could equally well have had 'cannot' instead of 'could not', but certainly not 'were unable to'. Indeed it could equally well have had neither, but simply the unmodalized 'so they were not married here'. The modalized forms 'could not/cannot have been' (or, in the case of (16), 'must have been') would be substituted for the unmodalized form only if there were a context of doubt or disagreement, calling for a reminder of argumentative cogency.

Take another example, one which I intend to display the commonplace nature of these expression-types. Someone looks out of a car window and says

(21) We must be getting near the city centre

or

(22) We can't be far from the city centre.

It would be a very feeble joke if his companion were to reply to (21), 'What do you mean? What's forcing us?', or to (22), 'What's stopping us?', since it is perfectly obvious that the first man is reporting an inference.

(ii) But is it an inference from premisses including one which contains a substantial impossibility? What could this be? According to the precedent of (M8), it would have to be

(23) It is impossible that (both) there is increased traffic density, more intersections, buildings, factories, . . . , and we are not nearer the city centre.

But who would recognize such a premiss as something disinterred from his actual thinking? What happens is surely that the inference goes straight from premisses about traffic density, etc., to the (unmodalized) conclusion. Of course such inferences have to be warranted by further factors, and are of various degrees of cogency. But there is no reason to think that, whenever an inference is signalled by a modal term, it must reflect a modalized premiss. Rather, it reflects whatever degree of cogency the inference possesses. Indeed I shall venture a counter-conjecture to that of Mavrodes: that, when a modalized premiss *is* involved, then the conclusion may be *doubly* modalized; and this may make possible a better account of (M2).

First, however, I want to scrutinize Mavrodes' claim, with which I began, that (M2) is 'neutral' with respect to the alleged impossibility of altering the past. This is supposed to be paralleled by the neutrality of (M6) with respect to the question of a wedding barrier. Now, although there was no wedding barrier of one kind, yet on Mavrodes' own account there was a barrier of another kind: the impossibility 'growing out of' bureaucracy, and described as one 'of some substance'. And this certainly sounds like a barrier that could be encountered, by someone, say, wishing to get married without leaving a record. Since this bureaucratic barrier is explicitly invoked in the analysis of the modality of (M6), it is hard to see how 'neutrality' is preserved. Likewise, then, for (M2): the 'substantial' impossibility invoked in a suppressed premiss must be one 'growing out of' – what? What else than the relevant facts about time and action? So how can (M2) be 'neutral' with respect to those?

This leads me to suggest that (M2) might be better analysed in terms of *double* modalization. First, consider a variant of (M2):

(24) She got married last week, so her father must be angry

which is the counterpart of a 'cannot' statement:

(25) She got married last week, so her father can't be pleased.

In both cases, the modal terms reflect conclusory force; the proper conclusions are, respectively, 'is angry' and 'isn't pleased'. But in addition we need to know what sort of premisses we are dealing with. In particular, do barriers occur in them? Whatever we may think of the bureaucratic barrier invoked in (M6) – and I, for one, would prefer to assimilate this example to my own 'no barrier' city traffic example – there can be no doubt that (25) invokes no barrier, whereas (M2) does. We may allow Mavrodes the possibility that it *may* not be absolutely certain that we cannot prevent what is done (though I think it *is* certain); what is undeniable is that we do say (and believe) that there are things we cannot do. When such a statement is, in addition, suitably embedded in an inferential context, as it is in (M2), we can have a double layer of modality – substance *and* shadow.

Unfortunately, although this analysis in terms of double modalization *could* be a correct analysis of (M2), I do not think that it is, so that Mavrodes is not even partially right. For if it were, 'he cannot' would have to elide 'it cannot be the case that he can'. But it does not have to; and a more natural construction is surely the normal one, 'it is not the case that he can' – an assertion unmodalized with respect to argumentative cogency, but modalized with respect to *de re* impossibility.

*The University, St Andrews,*
*Fife KY16 9AL*

© BERNARD MAYO 1983

**RATIO** edited by MARTIN HOLLIS

**Volume XXV (1983)** **June**

| | |
|---|---|
| CHRIS MORTENSEN and GRAHAM NERLICH | Spacetime and Handedness |
| C. DYKE | The Question of Interpretation in Economics |
| HANS-H. HOPPE | Is Research Based on Causal Scientific Principles possible in the Social Sciences? |
| CRAWFORD L. ELDER | Pragmatism and the World |
| STEPHEN READ | Quine's Private Language |
| HILLEL STEINER | Reason and Intuition in Ethics |
| JOACHIM SCHULTE | Wittgenstein and Conservatism |
| MURRAY MACBEATH | Mellor's Emeritus Headache |
| D. H. MELLOR | MacBeath's Soluble Aspirin |

*Subscriptions*

| | | | | |
|---|---|---|---|---|
| Institutions: | £22.95 (UK) | £28.20 (overseas) | $59.95 (USA) | $74.00 (Canada) |
| Individuals: | £15.50 (UK) | £19.00 (overseas) | $42.00 (USA) | $52.00 (Canada) |

*Orders to:* Journals Department, Basil Blackwell Publisher Ltd., 108 Cowley Road,
Oxford OX4 1JF, England.

The German edition of *Ratio* is published by FELIX MEINER VERLAG,
Richardstrasse 47, 2000 Hamburg 76, West Germany.

**Published by BASIL BLACKWELL, Oxford**